



HAL
open science

Bases de données cliniques et systèmes experts : méthodologie d'extraction de la connaissance

Alain Duhamel

► **To cite this version:**

Alain Duhamel. Bases de données cliniques et systèmes experts : méthodologie d'extraction de la connaissance. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1989. Français. NNT: . tel-00335614

HAL Id: tel-00335614

<https://theses.hal.science/tel-00335614>

Submitted on 30 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TU 33068

THÈSE

Présentée par

Alain DUHAMEL

Pour obtenir le titre de

Docteur de l'Université Joseph Fourier GRENOBLE I
spécialité GÉNIE BIOLOGIQUE et MÉDICAL

BASES DE DONNÉES CLINIQUES ET SYSTÈMES EXPERTS MÉTHODOLOGIE D'EXTRACTION DE LA CONNAISSANCE

Soutenue le 10 avril 1989

OOOO

Membres du jury

Président : Pr. J. DEMONGEOT

Rapporteurs : Pr. E. R. CARSON
Pr. M. FIESCHI

Examineurs : Pr. R. BEUSCART
Pr. G. COMYN

Thèse préparée au sein du laboratoire TIM3-IMAG

Je tiens tout d'abord à remercier Monsieur le Professeur Jacques DEMONGEOT, de l'Université Scientifique et Médicale de Grenoble, d'avoir accepté de diriger ce travail. Je voudrais lui exprimer ma profonde gratitude pour l'aide constante et les encouragements qui ont permis de le mener à bien.

Je remercie particulièrement Monsieur le Professeur CARSON, du Center for Measurement and Informations in Medicine (Londres), de m'avoir fait l'honneur de juger et de rapporter cette thèse.

Monsieur le Professeur Marius FIESCHI, de l'Université d'Aix-Marseille II, a accepté d'examiner et de rapporter ce travail. Je suis très honoré de le compter parmi mes juges.

Mes remerciements s'adressent également à Monsieur le Professeur Gérard COMYN, de l'Université des Sciences et Techniques de Lille, qui a accepté de juger cette thèse.

Je remercie vivement Monsieur le Professeur Régis BEUSCART, de l'Université de Lille II, pour la confiance qu'il m'a témoignée en m'associant à l'équipe du Centre d'Etudes et de Recherche en Informatique Médicale (CERIM) qu'il dirige. L'aide qu'il m'a apportée et les conseils avisés qu'il a bien voulu me donner au cours de ces trois années, ont contribué à la réalisation de cette thèse. Je lui suis très reconnaissant d'avoir accepté de juger ce travail.

Je tiens également à remercier Madame Claudine ROBERT, Maître de conférences à l'Université Scientifique et Médicale de Grenoble, pour sa collaboration à la première partie de cette thèse et la justesse de ses conseils en matière d'Analyse des Données.

Je remercie particulièrement Laurent MOUSSU, Ingénieur Système au CERIM, pour l'intérêt qu'il porte à l'ensemble des travaux de recherche réalisés au laboratoire. Je lui suis très reconnaissant pour l'aide efficace qu'il m'a apportée au cours de la réalisation de ce travail.

Je voudrais également remercier toute l'équipe du CERIM pour sa collaboration et sa constante bonne humeur.

Je n'oublierai pas celles et ceux qui ont contribué à la réalisation matérielle de cette thèse et en particulier Madame NIREL et Monsieur BRICHE. Je leur adresse mes plus vifs remerciements.

TABLE DES MATIERES

INTRODUCTION GENERALE	p 2
Chapitre I : Méthodologie	7
1. Introduction	7
1.1. Limites à l'utilisation des Systèmes Experts (SEM)	7
1.2. Objectifs de l'étude	9
2. Etude de la faisabilité d'un Système Expert	13
3. Principaux problèmes liés à l'évaluation	15
4. Résumé des problèmes liés à la validation	17
5. Comparaison avec la méthodologie essais thérapeutiques	19
5.1. Etapes I et II	
5.2. Etape III	21
6. Amélioration des performances	24
6.1. Notations et Définitions	25
6.2. Interaction avec l'Expert	27
6.3. Amélioration automatique : SEEK2	29
7. Méthodologie de validation et d'amélioration	34
7.1. Développement SEM-BD	34
7.2. Evaluation du SEM	35
7.3. Analyse des résultats	35
7.4. Etude des erreurs spécifiques	35
7.5. Comparaison règles/observations BD	37
7.6. Justification de la représentation (2 niveaux)	42
7.7. Retour aux experts	44
8. Faut-il rajouter des scores discriminants	45
9. Evaluer sur un échantillon test	46
10. Application des réseaux Bayésiens	47
11. Evaluation sur site	47
Chapitre II : Etude de faisabilité d'un SEM	50
1. Introduction	50
2. Le problème médical	51
3. Vers un système d'aide au diagnostic	53
3.1. Définition des connaissances	54
3.2. La base de cas	58
4. L'étude de faisabilité	59
4.1. Analyse univariée	59
4.2. Analyses multifactorielles	60
4.2.1. Analyse globale	61
4.2.2. Groupes sémantiques	65
4.2.3. Analyse discriminante	69
5. Conclusion de l'étude de faisabilité	71
6. Conclusion du chapitre	75

Chapitre III-A : Présentation du Système Expert Septicémie 78

1. Introduction	78
2. Le problème médical	79
3. La base de données	85
3.1. Structure du dossier	87
3.2. Recrutement	90
3.3. Etudes épidémiologiques	94
4. Le Système Expert	96
4.1. Partition en 2 modules	96
4.2. Données d'entrée	98
4.3. Représentation des connaissances	99
4.4. Acquisition des connaissances	102
4.5. Expertise	104
4.6. Utilisation du Système	108
4.7. Caractéristiques de SAM	109
4.8. Interfaces utilisateurs	115
4.9. Différences avec Mycin	116
4.10. Evaluations initiales	117

Chapitre III-B : Evaluation de SES sur la base de données 122

1. Transfert Base de données-Système Expert	122
1.1. Faisabilité et objectifs	122
1.2. Structure de LINXES	124
1.3. Description des modules	125
1.3.1. LINK1	125
1.3.2. LINK2	128
1.3.3. LINK3	131
2. Evaluation de la base de connaissances originelle	135
2.1. Critère d'évaluation	136
2.2. Performances globales	138
2.3. Analyse des réponses	143
2.4. Conclusion de l'évaluation	147
2.4.1. Comparaison avec les études antérieures	147
2.4.2. Difficultés de discrimination	150
2.4.3. Possibilités d'amélioration	151
3. Etude des difficultés de discrimination	152
3.1. Groupes "Ecoli", "Proteus", "Klebsielle"	153
3.1.1. L'expertise	154
3.1.2. Méthode et Résultats	157
3.1.3. Conclusion	163
3.2. Germes à faible prévalence	166
3.2.1. L'expertise	168
3.2.2. Les cas observés	171
4. Conclusion	173

Chapitre IV : Stratégie d'amélioration des performances	176
1. Introduction	176
2. Etape I : Groupes GBSV	178
2.1. Définition	178
2.2. Construction de base GBSV (DIAGSES1)	185
2.3. Evaluation de DIAGSES1	187
3. Etape II : Construction et évaluation de DIAGSES2)	190
3.1. Construction de DIAGSES2	190
3.2. Evaluation de DIAGSES2	198
4. Etape III : Analyses bivariées des règles	201
4.1. Construction de DIAGSES3	201
4.1.1. Méthode	201
4.1.2. Résultats	207
4.2. Evaluation de DIAGSES3	210
5. Etape IV : Analyses multifactorielles	213
5.1. Méthodologie : Analyse factorielle discriminante (AFD)	213
5.2. Comment construire les scores	216
5.3. Exemple d'analyse	217
5.3.1. Analyse globale	218
5.3.2. Analyse discriminante pas à pas et 2è AFD	221
5.3.3. Construction de scores	224
5.3.4. Sémantique des scores	227
5.3.5. Seuils décisionnels	230
5.3.6. Résultats	234
5.4. Résultats globaux	235
6. Conclusion	239
Chapitre V : Application de la Méthodologie des réseaux Bayésiens	244
1. Présentation du réseau utilisé	244
1.1. Introduction	244
1.2. Fonctionnement	246
2. Application au diagnostic des GBSV	248
3. Résultats.	250
Conclusion générale	254

INTRODUCTION GENERALE

INTRODUCTION

Lorsqu'une nouvelle molécule dotée de propriétés pharmacologiques est mise au point, le délai qui sépare sa synthèse en laboratoire de la mise sur le marché couvre plusieurs années (de 8 à 15 ans en moyenne). L'autorisation de mise sur le marché délivrée par les instances administratives de contrôle n'est octroyée qu'après des essais thérapeutiques conduits de manière rigoureuse respectant les quatre phases classiques (LAPLANCHE, 1987) :

PHASE I : Recherche, sur quelques volontaires, de la dose maximale tolérée.

PHASE II : Détermination de l'activité de la substance chez quelques sujets.

PHASE III : Essais comparatifs (à un placebo ou à un autre traitement efficace)

PHASE IV : Etudes de Pharmacovigilance.

C'est WYATT [WYATT 87] qui, l'un des premiers, a souligné le parallélisme méthodologique entre validation de système d'aide à la décision médicale et essai thérapeutique d'une nouvelle molécule pharmaceutique. Cette comparaison a le mérite de proposer un canevas de travail en vue de l'évaluation de ces systèmes. Elle met aussi en lumière les difficultés méthodologiques et la complexité du problème auxquelles le chercheur est confronté lorsque, ayant développé un programme d'aide à la décision, il désire convaincre ses pairs de l'intérêt de son logiciel, de son aptitude à répondre à la question posée et de son exactitude. En fait, il se trouve alors en amont même de la Phase I, tout comme le chercheur de laboratoire qui vient de

synthétiser une nouvelle molécule dont il suppose les effets pharmacologiques.

Il est parfois nécessaire de démontrer que le problème est soluble et que le système proposé présente effectivement un intérêt majeur dans l'aide au diagnostic ou au traitement. Il apparaît donc utile, dans certains domaines, d'effectuer des études de faisabilité destinées à démontrer que la tâche entreprise n'est pas d'avance vouée à l'échec. Nous proposons (cf. chapitre II) une méthodologie d'analyse d'un tel problème illustrée sur l'exemple du diagnostic des douleurs coronariennes.

Il apparaît de plus en plus nécessaire aussi que des critères rigoureux d'évaluation permettent à l'utilisateur d'estimer les performances du système d'aide à la décision. Actuellement, comme l'a souligné WYATT, peu de ces systèmes peuvent se targuer d'avoir dépassé la Phase II de l'essai. A notre connaissance, aucun de ces systèmes n'est entré en Phase III d'essai comparatif. Autant le critère paraît évident au cours des essais thérapeutiques (guérison, amélioration clinique ou biologique), autant il apparaît difficile à définir pour les logiciels d'aide à la prise de décision médicale : s'agit-il d'une amélioration de la prise en charge du malade, d'une meilleure adéquation traitement-maladie, d'une optimisation de la démarche diagnostique ou thérapeutique ? La définition de critères adéquats permettant des "essais comparatifs" reste bien souvent délicate et imprécise.

Tout au plus peut-on estimer simuler la Phase III en laboratoire en comparant, sur une base de cas parfaitement documentée, les performances du système d'aide à la décision aux performances des

experts ou à celles de non experts, voire de novices. Les simulations peuvent parfois apparaître artificielles, mais elles permettent, sur le critère simple (bien classé, mal classé), d'évaluer en première intention la pertinence du système d'aide à la décision, voire de le comparer aux médecins.

Au cours de ce travail, nous nous attacherons à démontrer la valeur d'une méthodologie associant, pour la validation d'un système d'aide à la décision médicale,

- une grande base de cas, parfaitement documentée,
- des médecins tests,
- des médecins experts.

Nous verrons que la démarche qui consiste à retourner aux experts tout nouveau résultat d'une évaluation amène ceux-ci à modifier la base de connaissances d'une manière pertinente. Dans certains cas, elle permet aussi de faire prendre conscience aux experts que la nature de la connaissance qu'ils utilisent quotidiennement n'est pas du même ordre que les connaissances livresques ou bibliographiques qu'ils croient nécessaires de mettre en avant. Nous verrons ainsi que l'analyse de l'évaluation du Système Expert Septicémie a permis de dégager le nouveau concept du "groupe bactérien de sensibilité voisine" [MOUTON 88] qui guide en fait l'attitude thérapeutique face à une maladie infectieuse grave.

Bien sûr, si l'évaluation du système n'est pas parfaite, le désir de son concepteur sera de l'améliorer, afin d'en augmenter la pertinence. Dans le cadre de systèmes experts médicaux utilisant le formalisme des règles de production pondérées par des coefficients de vraisemblance, nous proposons une stratégie de mise à jour des règles

et de leurs coefficients. Cette méthodologie insiste, ici encore, sur l'étroite collaboration avec les experts du domaine. Elle sera illustrée par la démarche qui a été menée dans le cadre de la mise à jour du Système Expert Septicémie. Dans certains cas, l'association de scores diagnostiques, introduisant une connaissance statistique multidimensionnelle, peut s'avérer intéressante. La transposition d'une partie du savoir sous forme de réseau bayésien peut aussi apporter des solutions à certains problèmes.

Nous espérons que la diffusion à grande échelle des systèmes experts médicaux permettra, dans le cadre d'une collaboration experts-médecins et utilisateurs-informaticiens, de mieux définir les critères d'efficacité d'un système d'aide à la décision médicale. Il se peut, en effet, qu'en pratique courante, ces critères d'efficacité soient différents du (bien classé, mal classé) que nous avons utilisé dans ce travail. La multiplication des postes de travail et des réseaux informatiques permet d'espérer que cette phase II-IV pourra être envisagée dans un proche avenir.

CHAPITRE I : METHODOLOGIE

CHAPITRE I : METHODOLOGIE

1 INTRODUCTION

La connaissance médicale est caractérisée par une progression rapide et une sophistication croissante. Face à cette "explosion de l'information" [REGGIA 85], il devient de plus en plus difficile pour les cliniciens de disposer de toutes les connaissances nécessaires en vue d'une prise de décision optimale [GOLDMAN 88]. Une réponse partielle a été obtenue par la multiplication des spécialités. Cependant, cette expertise demeure rare, disséminée, et difficilement accessible [WEISS KULIKOWSKI 85]. En 1959, LEDLEY [LEDLEY 59] a défendu l'idée selon laquelle un système d'aide à la décision médicale pouvait aider (et non remplacer) le praticien non expert à optimiser sa démarche diagnostique ou thérapeutique. Depuis, de nombreux systèmes d'aide à la décision médicale ont été construits, n'utilisant d'abord que des méthodes algorithmiques, probabilistes (Théorème de Bayes) [WARNER 61] [DE DOMBAL 74] [FRYBACK 78] ou basés sur la Statistique Décisionnelle (Discrimination) [ROMEDER 73] [NAKACHE 80].

1.1 Ces systèmes ont cependant montré leurs limites [REGGIA 85]

Le processus de décision médicale fait appel à des données imprécises, incertaines, et souvent incomplètes. Dans de nombreux domaines, la décision diagnostique et thérapeutique repose sur une connaissance qui dépasse largement la théorie limitée au domaine. Cette décision prend en compte des connaissances de nature variée : connaissances heuristiques, empiriques, évolutives.

A partir de 1970, le développement de l'Intelligence Artificielle [RICH 85] [FEIGENBAUM 85] [BARR 85], qui permet d'analyser et de reproduire le processus décisionnel à partir d'une représentation symbolique des connaissances, a apporté une alternative [SZOLOVITS 82] [KULIKOWSKI 80]. Dans de nombreux domaines de la médecine, des systèmes d'aide à la décision utilisant la méthodologie des systèmes experts [GONDRAN 85] [FARRENY 86] ont été développés. Depuis le plus connu d'entre eux, MYCIN [SHORTLIFFE 84], jusqu'aux plus récents [JOLY 88] [FARGEAS 88] [BEUSCART 87A], ces systèmes ont démontré leur capacité à résoudre des problèmes dans des domaines précis, mais complexes.

Pourtant, paradoxalement, ils demeurent peu utilisés. Selon REGGIA, cela est dû à quatre causes principales :

- a. L'absence d'importantes bases de données développées parallèlement au Système Expert Médical (SEM) : un système expert ne pourra être largement diffusé tant que la pertinence de ses décisions n'aura pas été vérifiée sur un grand nombre de cas cliniques.
- b. La portabilité : il ne s'agit pas de la portabilité physique de ces systèmes, mais de leur capacité à demeurer performants dans un environnement extérieur à leur site de développement. L'expertise peut en effet prendre en compte des paramètres locaux tels qu'une école d'attitude thérapeutique, ou s'adresser à des populations, utilisateurs ou patients, dont le profil dépend étroitement de critères géographiques.

- c. La résistance du corps médical : les médecins ne sont pas, actuellement, réellement convaincus de l'utilité de tels systèmes. Ils peuvent les ressentir comme une atteinte à leur intégrité professionnelle, la confusion ayant été parfois entretenue dans la définition du rôle clinique de ces systèmes pourtant destinés à aider le praticien et non à le remplacer. Les médecins considèrent par ailleurs qu'il n'existe pas de preuve concrète de leur utilité. Ils peuvent les juger rébarbatifs, coûteux, d'accès difficile. Trop souvent ces logiciels sont développés en laboratoire, sans consulter les utilisateurs potentiels. Des études ont montré que l'ergonomie ne devait pas être négligée dans la réalisation des systèmes experts, la qualité de l'interface Homme-Machine étant un facteur important pour leur acceptabilité [FOX 77], [GREMY 87].
- Ces préjugés défavorables semblent cependant évoluer, comme en témoigne le rapport de l'assemblée générale du Conseil National de l'Ordre des Médecins (Mai 88) [COM 88].
- d. La maintenance : la connaissance médicale évolue très rapidement et cette situation peut rendre un système expert rapidement obsolète.

1.2 Objectif de l'étude

L'objectif de notre étude est de montrer comment les bases de données cliniques peuvent apporter une aide à la validation des systèmes experts.

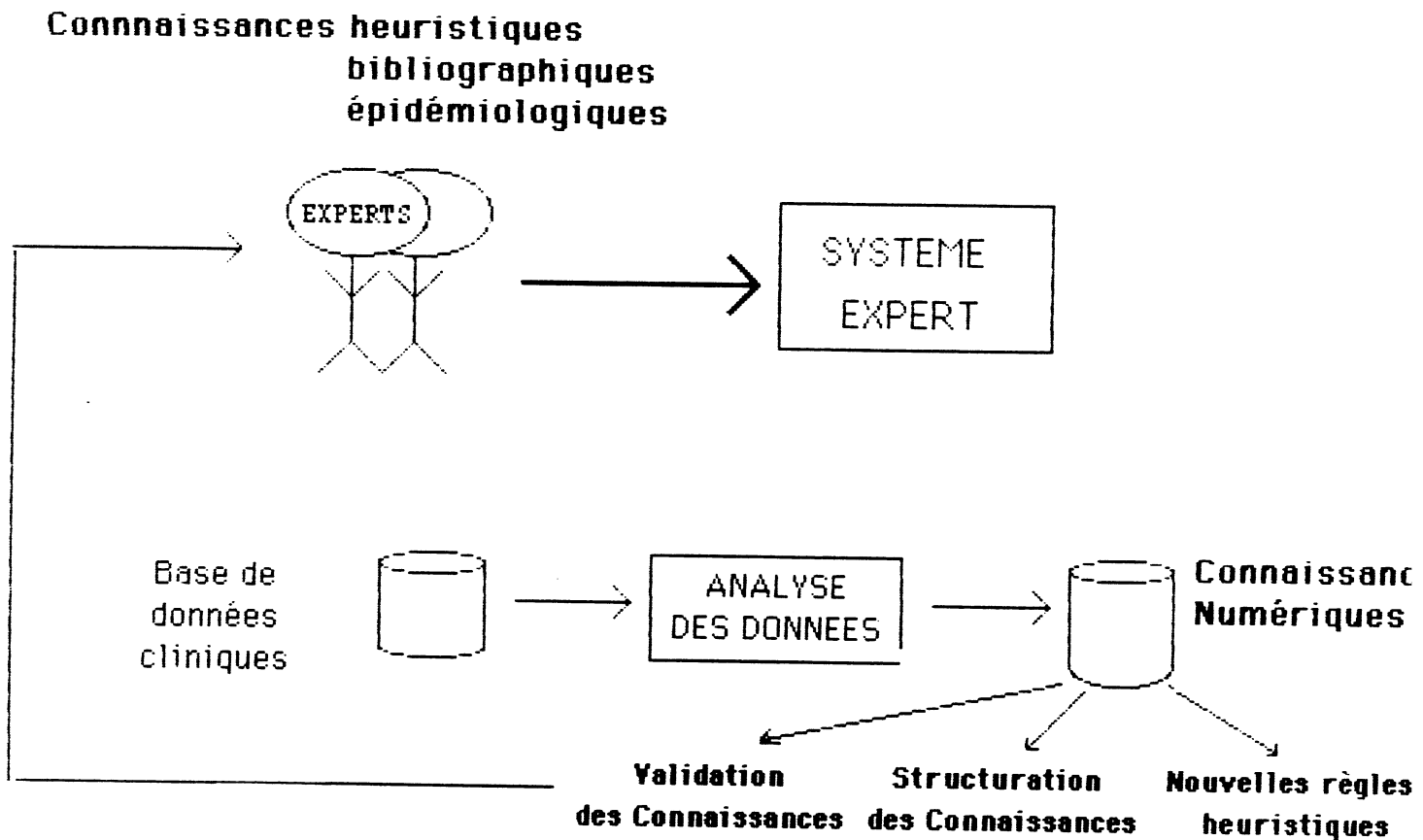


FIGURE 1.I

Insistons d'emblée sur plusieurs points qui sont à la base de notre démarche (FIGURE 1.I) :

- a. Notre étude concerne les domaines de la médecine pour lesquels il n'existe pas de théorie unificatrice, réunissant des connaissances imprécises, incertaines, empiriques, heuristiques. Dans ces domaines, l'expertise est nécessaire ; nous pensons qu'il n'est pas possible actuellement de modifier (a fortiori de créer), à partir d'une base de données clinique, un système expert sans l'accord des experts. Une base de cas, aussi importante soit-elle, ne peut pas, dans ces domaines, recouvrir toutes les occurrences possibles. De plus, on ne peut jamais garantir sa

représentativité. Donnons-en deux exemples :

- dans notre domaine d'application, l'infectiologie, nous avons développé une base de données multicentrique (19 centres français répartis sur l'ensemble du territoire) qui est actuellement l'une des plus importantes au monde (1500 septicémies) [PORTIER 87]. Un patient qui souffre d'une septicémie et qui présente une gangrène gazeuse a **sûrement** une infection à Anaérobie. Il s'agit d'une connaissance d'expert. Or, dans notre base de cas, aucun patient ne présentait ce symptôme. Cette connaissance n'aurait donc pas pu être découverte, dans une telle base.

- Notre recrutement, pourtant très important et multicentrique présente des caractéristiques particulières : notre population est généralement âgée, nos septicémies sont plus souvent acquises en service de médecine qu'en service de chirurgie... (cf. chapitre III A, paragraphe 3.2). Que dire de la portabilité d'une expertise déduite ou modifiée, sans l'accord des experts, à partir de cette base de cas ?

b. Nous nous plaçons donc en dehors de l'apprentissage tel que le définit [KODRATOFF 88] : "description d'un système incrémental qui, partant d'une connaissance quasi-nulle, améliore progressivement cette connaissance sur la base d'expériences qu'il éprouve". Notre objectif principal est de décrire une méthodologie qui permet d'améliorer la pertinence des décisions d'une base de connaissances déjà élaborée.

c. Dans ce cadre, nous considérons que les bases de données cliniques regroupent des "connaissances numériques" qui ne peuvent être ignorées dans le développement d'un système expert. Nous estimons que les connaissances symboliques des experts et les connaissances

numériques qui peuvent être extraites des bases de données cliniques par les techniques de l'analyse des données (analyses bivariées et multifactorielles), sont complémentaires [DUHAMEL 88c] [ROBERT 86] [RIALLE 87] :

c1) La connaissance n'est pas unique, les bases de données cliniques peuvent fournir un autre point de vue très fructueux. Ces connaissances numériques peuvent apporter un renouvellement dans le processus décisionnel des experts [GREMY 87]. Donnons-en un exemple. Face à une septicémie, il semblait nécessaire, pour instituer une antibiothérapie de première intention efficace, de déterminer le **germe** responsable de l'infection. Notre étude (chapitre IV, paragraphe 2.1) a montré que non seulement cette démarche n'était pas nécessaire, mais qu'elle n'était pas souhaitable. Avec les experts, nous avons montré que l'antibiothérapie de première intention efficace ne dépendait pas d'un germe précis, mais d'un concept plus large regroupant des germes pouvant être traité de la même manière, et la notion de Groupe Bactérien de Sensibilité Voisine a été introduite [MOUTON 88] [DUHAMEL 88a].

c2) Les connaissances symboliques des experts doivent être structurées de manière optimale. Le raisonnement des experts est très souvent un raisonnement sous incertitude. Face à ces problèmes, comme nous le verrons dans la suite de notre exposé, les bases de données peuvent apporter une aide fructueuse [DUHAMEL 88b].

d. La connaissance médicale évolue très rapidement. Une validation n'est donc jamais définitive [MANUEL 85], [FIESCHI 84a]. L'analyse des bases de données cliniques apporte une aide dans la

surveillance de cette évolution et dans la modification rendue nécessaire de la base de connaissances [BEUSCART 88a].

2 ETUDE DE FAISABILITE D'UN SYSTEME EXPERT

Avant d'envisager de construire une base de connaissances destinée à la discrimination, par exemple pour une tâche diagnostique, il est nécessaire de s'assurer que cette discrimination est possible. A partir d'un certain nombre de faits connus (Fi), il faut pouvoir discriminer entre des diagnostics possibles (Dj) pour en éliciter un ou plusieurs qui seront les plus probables.

Pour bâtir de telles bases de connaissances, il semble logique de se référer aux experts du domaine. Ceux-ci ont à leur disposition un savoir (éventuellement transmissible sous forme de règles de production) et une expérience ; ils peuvent aussi avoir recours à des "scores" de gravité qui les aident à établir le diagnostic ou le pronostic correspondant à un patient.

De même qu'un biologiste qui met en pratique un dosage, contrôle, sur la population dont il a la charge, la validité des normes qui lui sont fournies par le laboratoire, il apparaît indispensable que le "savoir" des experts soit vérifié à l'aide d'outils spécifiques, afin de s'assurer de sa qualité en termes de discrimination [ROBERT 88a] [COMBY 88].

On peut ainsi résumer notre démarche de la façon suivante :

l'expertise est limitée à la donnée de l'ensemble des évidences (Fi) et des conclusions (Dj). Les évidences Fi sont des signes ou syndromes que les experts jugent a priori pertinents pour déduire les conclusions Dj.

On dispose par ailleurs d'une base de cas W pour laquelle chaque cas est parfaitement documenté sur Fi et Dj. Il s'agit d'abord d'analyser le contenu de cette connaissance (1). Il s'agit ensuite d'évaluer la pertinence de cette connaissance vis-à-vis du problème de classification défini par (Fi, Dj) (2). Il s'agit enfin, en cas de réponse positive au (2), d'aider les experts à construire des règles pour développer un système expert.

Dans ce cadre, nous proposons d'utiliser les techniques de la Statistique descriptive multidimensionnelle (analyse en composantes principales) et de la statistique décisionnelle multidimensionnelle (analyse factorielle discriminante), afin de s'assurer que les items (Fi) permettent bien de discriminer les groupes diagnostiques (Dj) [FOUCARD 85], [BENZECRI 81], [SAPORTA 78], [LEBART 77], [ROBERT 89].

Dans le cas où l'on obtient une bonne discrimination entre les (Dj), cette étape peut être utile pour mettre en lumière les items les plus significatifs et les éventuelles associations de signes et de syndromes. Cette étape devra être confirmée par une concertation avec les médecins experts.

Dans le cas contraire, si l'on ne parvient pas à obtenir des discriminations satisfaisantes, il est très probable que l'on ne puisse pas structurer de base de connaissances permettant de résoudre le problème de diagnostic entre les (Dj). Cela signifie en effet qu'il n'est pas possible de calculer des fonctions linéaires des

variables retenues, qu'elles soient binaires (0,1) ou continues, pertinentes vis-à-vis du problème de classification.

Nous illustrerons au chapitre II cette problématique, en mettant en évidence la difficulté de réaliser un système expert dans le domaine de l'aide au diagnostic en Cardiologie (alors qu'il s'agit là d'un problème de premier plan en terme de santé publique). Cette constatation s'oppose à l'apparente simplicité des différents scores cliniques proposés par des "experts", dont nous montrons l'inefficacité.

3 PRINCIPAUX PROBLEMES LIES A L'EVALUATION

Nous avons parlé d'amélioration des performances d'un système expert. Qu'entend-on par performances d'un système expert, et comment les juger ? Ceci nous amène à étudier les principaux problèmes qui concernent l'évaluation des Systèmes Experts Médicaux (SEM).

De nombreux auteurs ont insisté sur la nécessité d'évaluer, par une méthodologie rigoureuse, les performances d'un SEM [BONNET 86], [SZOLOVITS 82], [SPIEGELHALTER 83], [FIESCHI 84a]. L'évaluation des SEM est une condition nécessaire à leur diffusion. Ces systèmes étant destinés à jouer un rôle clinique, il est de plus impératif d'évaluer la qualité de leurs décisions pour des raisons d'éthique.

Il existe cependant un grand nombre de critères de jugement. REGGIA indique trois points qu'il considère comme essentiels, s'agissant d'un système d'aide à la décision médicale :

- a. La justesse "absolue" : la réponse du SEM est-elle correcte ?
- b. La justesse "relative" : est-il plus performant que les utilisateurs potentiels ? Son utilisation est-elle bénéfique en terme d'économie de santé ?
- c. La portabilité.

 Pour une évaluation complète et s'agissant d'un système basé sur la connaissance, d'autres facteurs doivent cependant être pris en compte [BONNET 86], [FIESCH 84a] :

- d. Dans quelle mesure le SEM est-il capable de justifier ses réponses ? Deux raisons principales peuvent être avancées :

 d1) les cliniciens accepteront difficilement d'utiliser un système qui ressemblerait à une "boîte noire". TEACH [TEACH 81], en étudiant le comportement de cliniciens qui utilisaient des SEM, a montré que la qualité et la quantité des explications fournies par ces systèmes étaient des paramètres importants pour qu'ils soient acceptés.

 d2) La justification des réponses peut aider les experts à corriger des erreurs lors de la phase de mise au point de la base de connaissances [DAVIS 79].

- e. La sensibilité des réponses du SEM aux données d'entrée : si la pertinence des réponses du système expert dépend trop de la qualité et de la quantité des données qu'on lui fournit, il sera difficilement accepté [LANDEAU 87], [MANUEL 85] [BEUSCART 87a].

- f. La justesse des conclusions intermédiaires : YU [YU 79] a souligné que l'évaluation des réponses du Système ne pouvait se limiter à la seule vérification de ses conclusions finales : l'utilisateur acceptera d'autant plus facilement d'utiliser le SEM s'il retrouve un cheminement correspondant à un raisonnement cohérent.
- g. La facilité de mise à jour de la base de connaissances.
- h. La satisfaction des utilisateurs finaux.
- i. Le coût matériel, logiciel, les temps de réponses...

4 EN RESUME

Une validation complète d'un SEM est donc un problème complexe et ne peut être réalisée que par étapes.

Nous sommes conscients de la nécessité de prendre en compte l'ensemble des critères précédents. Les performances du SEM relatives à certains de ces facteurs dépendent cependant étroitement de la phase de développement : ainsi, le caractère explicatif du Système et la modularité de la base de connaissances sont liés au choix du formalisme de représentation des connaissances [FERRAND 84] [SWARTOUT 81].

La validation du SEM en terme de justesse "absolue" et "relative" est une condition nécessaire à sa diffusion. La mesure de la justesse "absolue" permet de disposer d'une référence pour l'amélioration de la qualité des décisions du SEM. Nous allons examiner dans quelle mesure il est possible de quantifier ces facteurs.

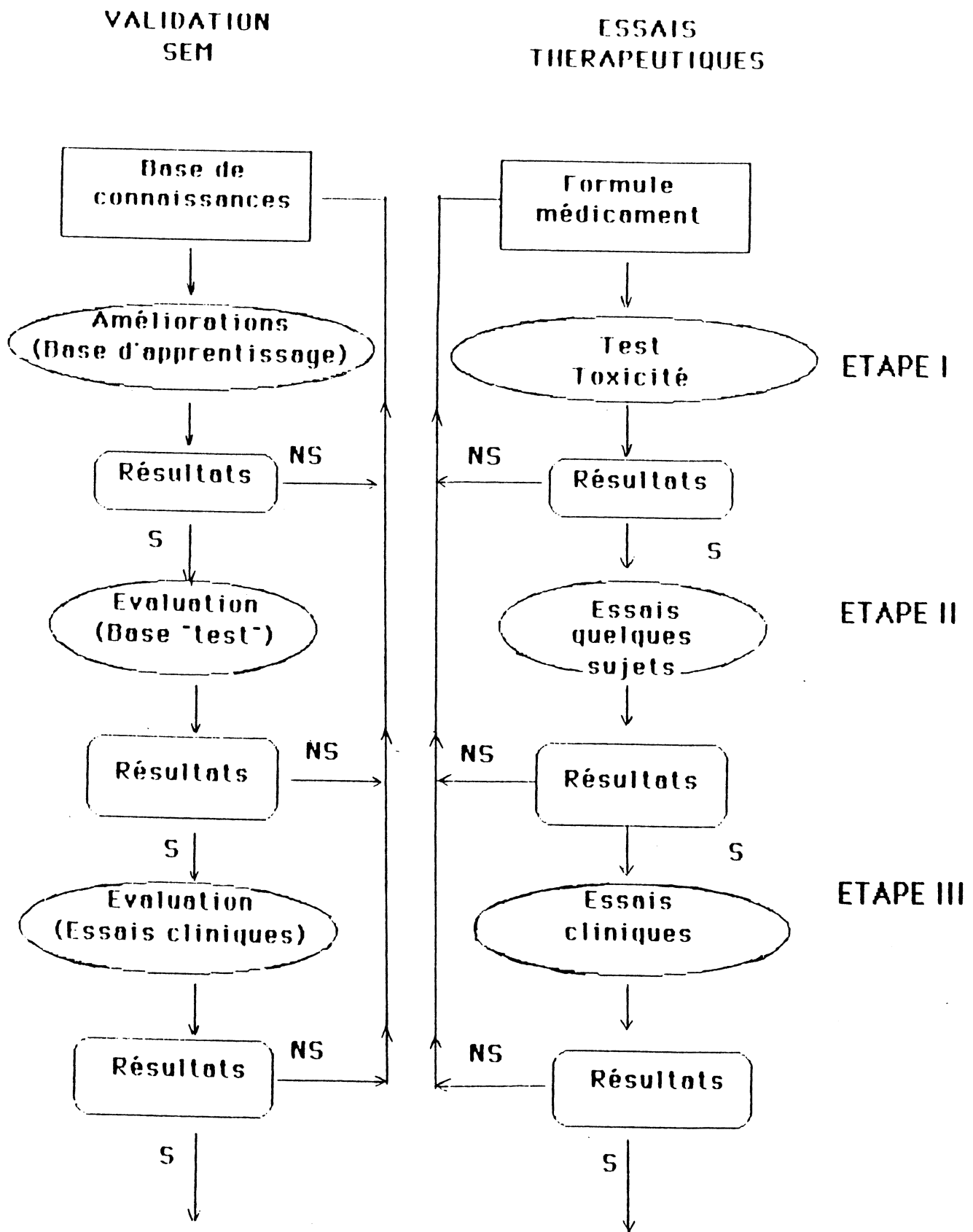


FIGURE 5 - 1

5 **COMPARAISON AVEC LE DEVELOPPEMENT D'UN MEDICAMENT**

WYATT [WYATT 87] a comparé les différentes étapes permettant d'aboutir à la validation globale d'un SEM, et donc en particulier à la validation en terme de justesse "absolue" et "relative", à celles réalisées pour le développement d'un nouveau médicament (FIGURE 5.I). On distingue trois étapes :

- a. Etape I : améliorations successives de la base de connaissances sur un échantillon d'apprentissage, par référence à la justesse "absolue" (sur dossiers).
- b. Etape II : Evaluation de la justesse "absolue" des réponses du Système sur un échantillon test (sur dossiers), distinct bien entendu de l'échantillon d'apprentissage.
- c. Etape III : Evaluation prospective en condition d'utilisation clinique (justesse "relative") : comparaison entre les réponses d'un médecin seul sur un premier groupe "placebo" et celles d'un médecin, aidé du SEM, sur un second groupe distinct "traité".

5.1 **Etapes I et II**

Comment quantifier la justesse "absolue" du Système ? En l'absence de base de données, il est nécessaire d'obtenir une base de cas pour l'évaluation. Ces dossiers devront être sélectionnés d'une manière rigoureuse. Nous renvoyons à [WYATT 87] et [MANUEL 85] pour une discussion de ce problème. Les dossiers participant aux étapes I et II devront bien sûr être distincts.

Le rôle clinique du Système étant défini, l'évaluation de la justesse "absolue" du Système nécessite la détermination de la réponse correcte ou "Gold Standard" de GASCHNIG [GASCHNIG 83] pour chaque dossier participant à l'évaluation.

Pour déterminer de manière certaine la réponse correcte, il faut pouvoir se référer à un test objectif, indiscutable. Dans de nombreux domaines d'application des SEM, ce test objectif n'existe pas : pour les systèmes d'aide à la thérapeutique, par exemple, il est difficile, devant un cas clinique, de déterminer "le" traitement théorique; la vérification d'un diagnostic peut nécessiter des examens complémentaires invasifs ou coûteux.

Dans de telles situations, on considère que le "Gold Standard" peut être obtenu par un consensus de plusieurs experts du domaine [YU 79], [FIESCHI 84a]. L'avis d'un seul expert est en effet bien souvent insuffisant. Face à une même pathologie, les réponses des experts peuvent diverger : il peut exister, par exemple, des différences d'appréciation dans le choix des traitements.

Les experts participant à la détermination de la réponse correcte doivent être indépendants entre eux et vis-à-vis du SEM : en prenant les avis des experts qui ont développés la base de connaissances, on introduirait un biais très favorable au SEM [MANUEL 85]. Par ailleurs, la reproductibilité des avis des experts doit être établie [FOX 85].

Les dossiers retenus pour l'évaluation de la justesse "absolue" seront ceux pour lesquels la réponse correcte a pu être établie.

Lorsque ce "Gold Standard" est ainsi déterminé ou lorsqu'il est connu de manière certaine (par exemple, dans le cas d'un diagnostic déterminé par des examens complémentaires), on peut obtenir un coefficient d'efficacité du système expert, en comparant ses réponses aux réponses correctes pour chaque dossier participant à l'évaluation. Ce coefficient peut servir de référence pour améliorer les performances du SEM lors de l'étape I et vérifier sa pertinence lors de l'étape II [GINSBERG 88], [WEISS-KULIKOWSKI 85], [DUHAMEL 88c].

Ce coefficient donne un point de vue. Il est utile pour améliorer les performances mais il ne règle cependant pas tous les problèmes [REGGIA 85], [SHAPIRO 77]. Supposons en effet que le Système fournisse sa réponse sous la forme de plusieurs hypothèses pondérées [SHORTLIFFE 75], et considérons, pour un même cas clinique, les deux réponses :

H1 est le "Gold Standard".

	Réponse I	Réponse II
H1	vraisemblance 0,8	H1 vraisemblance 0,8
H2	vraisemblance 0,5	H2 vraisemblance 0,79
		H3 vraisemblance 0,78

Il est clair que la réponse I est la meilleure. La situation est plus grave si, la première hypothèse citée étant correcte, la deuxième s'avère gravement fausse. Il faut donc, dans le cas de tels systèmes, faire intervenir la capacité de jugement d'un expert [SOULA 87].

5.2 Etape III

Dans cette étude, on cherche en premier lieu à montrer que la décision clinique d'un médecin aidé d'un système expert est meilleure que celle d'un médecin non aidé.

Comme WYATT, on peut considérer ce problème comme la réalisation d'un test de TURING [SPIEGELHALTER 83] [DELAHAYE 87] [RICH 85] (FIGURE 5.II).

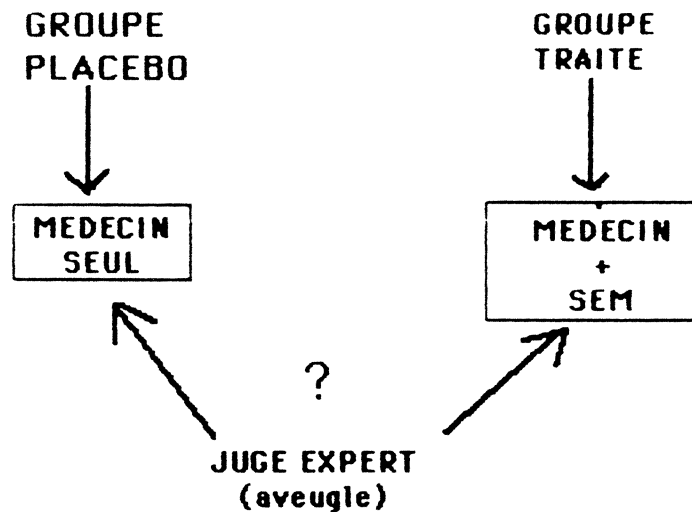


FIGURE 5.II

Il est alors indispensable de prendre en compte les points suivants [MANUEL 85], [QUAGLINI 87] :

a. Les groupes "placebo" et "traité" doivent être indépendants et comparables.

b. Méthodologie en "aveugle" : à aucun moment, l'expert ne doit savoir qui il juge.

c. Etude de concordance : exprimer les résultats sous la forme d'un pourcentage de réponses identiques entre chaque médecin et l'expert est insuffisant, dans la mesure où une identité de point de vue peut être le fruit du hasard. COHEN [COHEN 60] a introduit le coefficient de KAPPA qui permet de tester si la concordance observée est significativement différente de la concordance due au hasard :

$$K = \frac{P_o - P_c}{1 - P_c}$$
$$P_o = \% \text{ de réponses identiques.}$$
$$P_c = \% \text{ de réponses identiques dues au hasard.}$$

Pour un échantillon suffisamment grand ($N > 30$),
K suit une loi normale.

La démarche correspondant à cette étape III se heurte cependant selon nous aux problèmes suivants :

- Pb 1 : Un système expert n'est pas un médicament : il n'a pas d'efficacité en lui-même. Il est destiné à suppléer aux défaillances du médecin. Comme le souligne [SPIEGELHALTER 83], le SEM compte sur le clinicien utilisateur pour être efficace.
- Pb 2 : La méthodologie décrite à cette étape par WYATT nécessite un préalable : Les cliniciens participant à l'évaluation doivent avoir des niveaux de compétence comparables. Ceci suppose donc que l'on puisse comparer les connaissances de médecins en exercice, ce qui nous semble peu réaliste eu égard au respect des principes éthiques et déontologiques. Pour se rapprocher d'une utilisation en grandeur réelle d'un SEM, il faudrait pouvoir réaliser ce type d'étude au niveau de plusieurs hôpitaux. Il serait alors possible d'évaluer réellement les conséquences de son utilisation d'une façon individuelle, mais aussi globale : la décision clinique du médecin aidé par le SEM est-elle plus pertinente pour le malade ? Quels sont les effets de l'utilisation du SEM en termes de nombre d'hospitalisations, de durée d'hospitalisation, de coût ? Ce qui suppose que le SEM soit déjà très largement diffusé.

Les problèmes liés à l'évaluation étudiés dans ce paragraphe montrent qu'il est très difficile de juger de la validité globale d'un SEM. Sans que cela signifie qu'il ne faille pas évaluer, nous sommes tentés de dire que les SEM vérifieront un jour "le principe de l'aspirine" : on ne saura peut-être d'une manière certaine s'ils sont utiles que le jour où ils seront très largement diffusés.

6 AMELIORATION DES PERFORMANCES

Nous supposerons dans ce paragraphe que l'on dispose d'une base de connaissances élaborée. Plus précisément, nous considérerons que toutes les connaissances symboliques des experts ont été transférées et représentées, dans un formalisme choisi, dans la base de connaissances : faits pertinents pour le domaine, modèles de raisonnement.

Plaçons nous dans le cadre des modèles de classification [WEISS-KULIKOWSKI 85]. Dans un modèle de classification, toutes les solutions possibles sont déterminées a priori. La tâche du Système consiste alors à affecter chaque cas clinique qui lui est soumis à une classe : par exemple, déterminer un diagnostic parmi une liste de diagnostics possibles. Dans le formalisme des systèmes de production [LAURIERE 82] [NILSSON 80], ces modèles gèrent, de manière synthétique :

- un ensemble de faits ou d'évidences possibles,
- un ensemble de classes possibles,
- un ensemble de règles liant les faits aux classes. Ces règles modélisent la stratégie de raisonnement des experts. La structure générale d'une règle de production est la suivante :

"Si Prémisse Alors Conclusion ",
où la prémisse décrit, dans un formalisme défini, un contexte clinique, et la conclusion, l'action à réaliser dans ce contexte. Ces règles de production peuvent être pondérées par des Coefficients de Vraisemblance (CV) qui expriment la confiance de l'expert en la règle [SHORTLIFFE 75] [GASCUEL 81].

6.1 Hypothèses, notations et définitions

Dans la suite de notre exposé :

6.1.1 Nous supposerons que l'on dispose d'une base de cas comportant N individus.

6.1.2 Nous désignerons par :

(Fi) (i= 1,K) l'ensemble des faits possibles,

(Di) (i= 1,L) l'ensemble des conclusions possibles,

(Ri) (i= 1,P) les règles de la base de connaissances,

(Pi) (i= 1,Q) les prémisses des règles, qui mettent en relation des faits Fi à l'aide des opérateurs logiques ET, NON.

6.1.3 Nous supposerons que le "Gold Standard" Gj est connu pour chacun des individus j de la base de cas.

Nous désignerons par G la variable qualitative "référence" : $G=G_j$ pour l'individu j; les modalités de G sont D1....DL.

6.1.4 De manière analogue C désigne la variable qualitative "Conclusion" du Système".

6.1.5 Définition de l'évaluation

Nous dirons que la base de connaissances est évaluée sur un sous-ensemble W de la base de cas :

a. si l'on dispose sur W du tableau de contingence G x C (**Référence**) X (**Conclusions SEM**). Ce tableau est une matrice de dimension LxL que nous appellerons Tw.

G \ C	D1	D2	...	Di	...	DL
D'1	n11	n12	...	n1i	...	n1L
D'2	n21	n22	...	n2i	...	n2L
.
D'j	nj1	nj2	...	nji	...	njL
.
D'L	nL1	nL2	...	nLi	...	nLL

Le nombre total d'individus de W "bien classés" par le SEM est donc la trace de Tw, c'est-à-dire la somme des éléments de la diagonale. Le but principal d'une amélioration est d'obtenir une matrice Tw la plus proche possible d'une matrice diagonale. Ce tableau permet, en outre, d'apprécier la qualité de la classification réalisée par le SEM. On dispose, en effet, pour chaque conclusion Di possible :

- du nombre de faux positifs (FPi) et faux négatifs (FNi),
- on connaît le nombre d'entre eux qui ont été affectés de manière erronée à telle ou telle classe.

b. si, pour chaque individu de W, les conclusions du Système sont archivées.

Dans la suite de notre exposé, W désignera l'échantillon

d'apprentissage.

6.1.6 Définition de l'amélioration

Nous considérerons qu'une stratégie S améliore les performances de la base de connaissances si :

- En utilisant la stratégie S, on modifie la base de connaissances : ajouts de règles, suppressions de règles, modifications de prémisses, modifications de coefficients de vraisemblance. Cette stratégie est développée en prenant en compte le sous-ensemble W de la base de cas précédemment défini.

- Les modifications induites par S sont sémantiquement correctes : elles ont été ratifiées par les experts du domaine.

- Désignons par V un sous-ensemble de la base de cas tel que $W \cap V = \emptyset$. V est constitué de cas nouveaux sur lesquels la base de connaissances initiale est évaluée. Soit Nbi le pourcentage de "bien classés" alors obtenu sur V.

- Après évaluation de la nouvelle base de connaissances sur V, on obtient un pourcentage de "bien classés" NBf tel que $NBf > Nbi$, et les résultats de cette évaluation sont jugés qualitativement corrects, après vérification par les experts.

6.2 Améliorations des performances : interaction avec l'expert

De nombreux travaux ont été consacrés au problème de l'acquisition des connaissances. Ils concernent en particulier la description de méthodologies permettant de mieux structurer les différentes phases d'élaboration d'une base de connaissances

[HAYWARD 87] ou la construction de systèmes d'aide à l'acquisition des connaissances destinées principalement aux cognitivistes comme AQUINAS [BOOSE 87]. D'autres recherches ont été réalisées pour développer des systèmes permettant de transférer et d'organiser les connaissances, et de construire un système expert par interaction directe avec les experts du domaine : MORE [KAHN 85], TEIRESIAS [DAVIS 79], MOLE [ESHELMAN 86].

Concernant la phase d'amélioration des performances, certains systèmes valorisent l'intervention des experts : TEIRESIAS [DAVIS 79], MOLE [ESHELMAN 86]. Ces systèmes permettent de corriger les règles lorsque, pour un cas donné, la conclusion est erronée. Pour ce faire, ils sont capables de dérouler le raisonnement ayant abouti à une conclusion fautive, et de guider l'expert dans le choix des modifications à réaliser. Ils peuvent apporter une aide à la mise au point de la base de connaissances, mais comportent des limites :

a. en travaillant au "cas par cas", ces systèmes ne donnent pas la possibilité aux experts d'apprécier l'ensemble des problèmes comme le permet une évaluation telle que nous l'avons décrite.

b. Il n'y a pas de contrôle des modifications réalisées sur une base de cas, en termes de performance globale et de performances liées à chaque conclusion possible.

c. Une modification au "cas par cas" peut induire des biais dans la base de connaissances, en particulier si les modifications sont réalisées à partir de cas cliniques difficiles.

d. Aucune connaissance nouvelle n'est apportée à l'expert pour l'aider à modifier la base de connaissances. Tout repose, comme lors

de la phase d'acquisition des connaissances, sur les connaissances symboliques des experts.

6.3 Stratégie d'amélioration automatique : SEEK2

POLITAKIS ([POLITAKIS 84] "SEEK") et GINSBERG ([GINSBERG 88] "SEEK2") ont proposé une méthodologie d'amélioration d'une base de connaissances à partir d'une base de cas, qu'ils ont appliquée aux systèmes à base de règles de production. Celle-ci est actuellement intégrée au shell du système expert "EXPERT" [WEISS 79].

6.3.1 La démarche employée

La démarche repose sur la contrainte de "conservatisme" : les améliorations recherchées doivent préserver au maximum la base de connaissances initiale. Nous reviendrons sur ce point. La méthodologie est basée sur les concepts de généralisation et de spécialisation [MITCHELL 82] :

Généralisation

Généraliser une règle signifie affaiblir la condition de déclenchement de la règle ou augmenter son coefficient de vraisemblance.

Spécialisation

Spécialiser une règle signifie au contraire renforcer sa condition de déclenchement ou diminuer son coefficient de vraisemblance :

EXEMPLES DE GENERALISATIONS

SI (E1 et E2 et E3) ALORS D1 -----> SI (E1 et E2) ALORS D1
 SI (E4) ALORS D2 <.6> -----> SI (E4) ALORS D2 <.8>

EXEMPLES SPECIALISATIONS

SI (E1 et E2) ALORS D1 -----> SI (E1 et E2 et E3) ALORS D1
 SI (E4) ALORS D2 <.9> -----> SI (E4) ALORS D2 <.7>

Un processus d'amélioration des performances tel que nous l'avons décrit doit nécessairement maximiser Nb (le pourcentage de "bien classés"). Pour ce faire, GINSBERG propose de travailler localement : soit Df une conclusion fixée. On désigne par Mf l'ensemble des individus de la base de cas W "mal classés" pour Df. Le cardinal de Mf est $CARD(Mf) = FPf + FNf$ (faux positifs + faux négatifs) ; maximiser les performances du Système pour Df est donc équivalent à minimiser $CARD(Mf)$. Dans ce cadre, une généralisation sera une tentative pour diminuer FNf, une spécialisation, une tentative pour diminuer FPf. Naturellement, une amélioration locale peut dégrader les performances globales du Système. SEEK2 recalcule donc Nb après chaque amélioration locale, pour en mesurer les conséquences globales.

Ainsi, le choix des modifications proposées (généralisation ou spécialisation) repose sur deux points :

a. les résultats des performances du Système sur une base de cas W : performances globales, performances pour chaque conclusion possible, réponses du Système pour chaque individu de la base de cas.

b. le "comportement statistique" des règles, évalué sur W.

6.3.2 Evaluation du comportement statistique d'une règle

Le "comportement statistique" d'une règle R est estimé grâce à plusieurs paramètres. Deux nombres entiers, "GEN(R)" et "SPEC(R)", sont à la base du processus d'amélioration. "GEN(R)" (respectivement "SPEC(R)") permet de quantifier, à partir des observations de W, l'aptitude d'une règle R à la généralisation (respectivement à la spécialisation). Ainsi, le système proposera la généralisation de R si, après calcul, on obtient $GEN(R) > SPEC(R)$. Examinons de quelle façon sont calculés ces coefficients :

GEN(R) et SPEC(R) sont initialisés à 0. En reprenant les notations précédentes, on désigne par Df une conclusion du Système, j un individu de Mf, et (Rf) l'ensemble des règles qui déduisent "Df". Pour toute règle R de (Rf), il est possible de déterminer pour l'individu j si :

- a. **La règle a été déclenchée à tort** : si elle avait été "bloquée", la bonne conclusion aurait été atteinte. R est alors candidate à la spécialisation et SPEC(R) est incrémenté de un.
- b. **La règle aurait dû être déclenchée** :
 - (1) Df est la bonne conclusion, mais la prémisse n'est vérifiée qu'en partie (respectivement le coefficient de vraisemblance de R est trop faible) ;
 - (2) si la règle avait été déclenchée (respectivement si on augmentait le coefficient de vraisemblance de R), le Système aurait conclu "Df" en première position.
 - (3) parmi toutes les règles vérifiant (1) et (2), R est celle qui nécessite le moins de modifications, en termes de nombre de prédicats conjonctifs à éliminer, pour être déclenchée. R est

candidate à la généralisation et GEN(R) est incrémenté de un. Outre GEN(R), le Système identifie alors le(s) fait(s) qu'il faut supprimer de la prémisse (respectivement l'augmentation du coefficient de vraisemblance) pour que la règle permette de déduire la conclusion correcte.

6.3.3 Principe de fonctionnement

Le Système utilise alors des heuristiques pour proposer la généralisation ou la spécialisation d'une règle, par exemple :

soit P la prémisse d'une règle R telle que $P=(F1 \text{ et } F2 \text{ et } \dots F_n)$; une heuristique possible est :

"SI GEN(R) > .SPEC (R) et SI F_k est le fait le moins souvent vérifié en prémisse de R, ALORS supprimer F_k".

- a. La généralisation : les heuristiques de généralisation proposent soit de supprimer des faits en prémisse, soit d'augmenter le coefficient de vraisemblance de la règle.
- b. La spécialisation : les heuristiques de spécialisation ne proposent pas d'ajout de faits en prémisse. La spécialisation est réalisée en diminuant le coefficient de vraisemblance de la règle.

6.3.4 Comparaison avec notre méthode

Cette procédure permet d'affiner les performances de la base de connaissances. Chaque cycle intègre dans la base de connaissances la modification qui maximise l'augmentation du pourcentage de "bien classés" Nb sur W. Un échantillon test V peut être défini, pour juger de la stabilité des résultats. L'expert reste le seul juge de la

validité des améliorations ainsi déterminées.

Cependant, le processus décrit par GINSBERG ne peut pas réellement découvrir et proposer des connaissances nouvelles. La stratégie d'amélioration employée est comparable à une optimisation de la base de règles sur l'échantillon W. Il n'est pas possible, avec cette démarche, d'apporter un renouvellement de la connaissance, d'intégrer dans la base de connaissances des nouvelles règles qui pourraient s'avérer intéressantes ou de supprimer des règles non pertinentes [DUHAMEL 88a].

L'analyse des données [SAPORTA 78] [FOUCARD 84] [BENZECRI 81] peut permettre d'extraire des bases de données cliniques des informations nouvelles, complémentaires des connaissances symboliques des experts [ROBERT 88c], [DUHAMEL 88c] [RIALLE 87]. Les techniques de l'analyse des données que nous avons utilisées ont été choisies en fonction de notre étude. Ce choix illustre cependant l'utilisation qui peut être faite de l'analyse des données en relation avec les systèmes experts.

Le paragraphe suivant décrit notre méthodologie d'aide à la validation des SEM. Cette méthodologie a été appliquée au Système Expert Septicémie [BEUSCART 87b]. Elle tente de prendre en compte les limitations à la diffusion des SEM relevées par REGGIA. L'un des objectifs principaux est d'améliorer le pouvoir décisionnel d'une base de connaissances. Dans ce cadre, la stratégie S d'amélioration des performances a été développée à partir d'analyses bivariées effectuées sur les règles d'inférence, d'analyses factorielles discriminantes et de la ROC analyse.

7 **METHODOLOGIE DE VALIDATION ET D'AMELIORATION DES PERFORMANCES
D'UN SEM**

7.1 Développement concomitant du SEM et d'une base de données (BD)

On cherchera à optimiser la cohérence entre la BD et le SEM de la manière suivante (chapitre III A. :

a. le SEM sera développé si possible à partir d'une expertise multiple et multicentrique [BEUCART 88a].

b. La BD sera alors implantée dans chaque centre. Une centralisation des dossiers sera prévue. Le SEM et la BD centralisée devront être interfacés de manière à rendre possible l'exécution du SEM sur un ensemble W choisi des dossiers de la base de cas, et donc son évaluation sur W (chapitre III B).

c. Le choix des items de la base de données est capital. Il devra être effectué en prenant en compte le problème de la validation du SEM. Plus précisément, ce choix devra comprendre l'ensemble des faits initiaux qu'il est possible de fournir au SEM, mais aussi l'ensemble des sous-butts du SEM de façon à rendre possible une évaluation du cheminement du raisonnement. Ce choix devra intégrer l'ensemble (Di) des conclusions du SEM : la variable qualitative G ("référence") devra être enregistrée pour chaque dossier.

d. Le nombre d'items de la BD ne devra pas être limité à une stricte transposition des faits et conclusions manipulés par le SEM : une probable évolution des connaissances manipulées par le Système devra être prise en compte.

7.2 Evaluation du SEM (chapitre III B)

Il faut d'abord déterminer un échantillon d'apprentissage W . W regroupe un ensemble de dossiers représentatifs de la BD (tirés au hasard). Puis évaluer le SEM sur W . On obtient ainsi le tableau de contingence $G \times C$ (T_w) et les conclusions du SEM pour chaque dossier. Enfin, étudier T_w : on dispose du coefficient global d'efficacité du SEM sur W (le pourcentage N_b de "bien classés"), du pourcentage de bonnes réponses du SEM pour chaque conclusion D_i , on peut donc apprécier la qualité des réponses du Système.

7.3 Analyse des résultats de l'évaluation avec les experts
(chapitre III B)

Il faut alors distinguer :

- a. les erreurs "admissibles" au niveau actuel de l'expertise : erreurs dues à la qualification de l'expert, à la nature de la décision médicale.
- b. les erreurs injustifiables intellectuellement.

7.4 Etude spécifique des erreurs "admissibles" (chapitre III B)

Désignons par D_1 , D_2 et D_3 , trois conclusions que les règles de la base de connaissances ne permettent pas de discriminer de manière satisfaisante. Les objectifs de cette étude sont de répondre aux questions suivantes :

- a. Peut-on distinguer ces conclusions par des combinaisons linéaires de faits? Peut-on rajouter certains types de conclusions?

Notre objectif est d'améliorer le pouvoir décisionnel du SEM. Il s'agit donc de déterminer s'il est possible de découvrir, à partir des observations de la BD, de nouvelles associations de faits, pertinentes pour distinguer les conclusions D1, D2, et D3.

Pour ce faire, nous proposons de réaliser une Analyse Factorielle Discriminante (AFD) sur les groupes W1 (sous-ensemble des individus de W tels que $G_j=D1$), W2 et W3. La sélection des variables participant à l'analyse se fera avec les experts.

Si les résultats de cette analyse ne permettent pas de distinguer les trois groupes, on passera à la question b. Si les résultats sont positifs, il faut successivement :

1) Construire des scores $S1=CL1(F_i)$ (combinaison linéaire des F_i) et $S2=CL2(F_i)$ à partir des variables les plus responsables de la discrimination (ces variables pourront être déterminées par une analyse discriminante "pas à pas" [ROMEDER 73]). Puis déterminer les seuils prédictifs de ces scores par la ROC analyse en calculant leur sensibilité et leur spécificité (chapitre IV.5).

2) Fournir les résultats aux experts. Etudier la sémantique de ces scores : un score, même s'il possède un bon pouvoir discriminant, ne sera pas accepté par le clinicien s'il ressemble à une "boîte noire". Pour expliquer la sémantique d'un score S1, les résultats fournis par l'AFD (valeurs des coefficients de CL1, corrélation $S1/F_i, \dots$) devront être confrontés aux règles de la base de connaissances (chapitre IV.5).

3) Avec les experts, modifier la base de connaissances : regrouper les conclusions D1, D2 et D3 en un sous-but. Intégrer

les règles associées aux scores [ROBERT 86], ayant le sous-but pour prémisses.

4) Evaluer la nouvelle base de connaissances.

- b. Les résultats de l'AFD ne permettent pas de distinguer les groupes W_j . Quelles sont les conséquences cliniques de ces proximités diagnostiques ?

Déterminer, avec l'aide des résultats fournis par l'AFD et en étudiant les règles de la base de connaissances, pour quelles raisons les conclusions D1, D2, et D3 ne sont pas distinguées par le SEM. Pour ce faire, comparer les règles qui déduisent les groupes D1 D2 et D3; sur le graphique fourni par l'AFD, peut-on confirmer ou déterminer des profils cliniques : quelles sont les caractéristiques cliniques des individus proches ? Existe-t-il des groupes bien séparés (différents de W1, W2, et W3) ? Quelles sont leurs caractéristiques cliniques ? (chapitre III B, Paragraphe 3)

Evaluer les conséquences de ces phénomènes de proximité vis-à-vis du processus de décision dans lequel s'intègre le SEM : peut-on changer le type de conclusion de l'expertise ? En cas de réponse positive, modifier et réévaluer la base de connaissances (chapitre IV, Paragraphe 2).

7.5 Comparaison des règles aux observations de la BD

Différentes méthodes numériques [BHATNAGAR 86] sont actuellement utilisées dans les SEM pour prendre en compte l'incertitude de l'information : théorie des possibilités [PRADE 85] (PROTIS

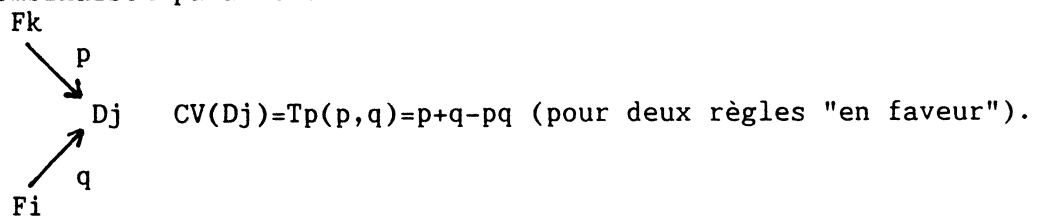
[SOULA 82]), théorie de l'évidence [SHAFER 76] (GENDIAG [GOUVERNET 85]), coefficients de vraisemblance [SHORTLIFFE 75] (MYCIN, SAM [GASCUEL 81]).

Plaçons-nous ici dans le cadre des règles de production affectées de coefficients de vraisemblance. Lors de la phase d'acquisition des connaissances, le coefficient de vraisemblance d'une règle R est obtenu de manière classique en demandant à l'expert d'indiquer, à l'aide d'une valeur comprise entre 0 et 1, le degré de confiance (respectivement de défiance) il accorde à la conclusion Dj en présence du fait Fi. Lorsque la conclusion Dj est positive, il s'agit d'un coefficient "en faveur" de Dj (CV0), dans le cas contraire d'un coefficient en défaveur de Dj (CVN). Ainsi la règle R peut s'exprimer sous la forme :

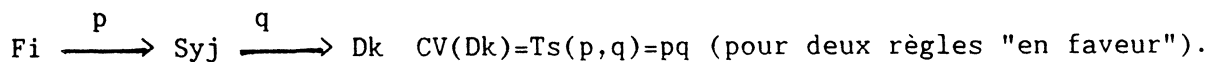
"SI Fait Fi alors Conclusion Dj <CV0> (respectivement NON Dj <CVN>)"

Des formules de combinaison (parallèle Tp et séquentielle Ts) permettent de propager le raisonnement incertain à travers le réseau d'inférence :

Combinaison parallèle



Combinaison séquentielle



SHORTLIFFE [SHORTLIFFE 75] a proposé une définition probabiliste des coefficients de vraisemblance (CV_P). Pour la règle précédente R :

$$(CV_P) \left\{ \begin{array}{l} CV0 = 1 \text{ si } P(Dj) = 1, \\ \quad = [P(Dj/Fi) - P(Dj)] / [1 - P(Dj)] \text{ si } P(Dj/Fi) > P(Dj), \\ \quad = 0 \text{ sinon.} \\ \\ CVN = 1 \text{ si } P(Dj) = 0, \\ \quad = [P(Dj) - P(Dj/Fi)] / P(Dj) \text{ si } P(Dj/Fi) < P(Dj), \\ \quad = 0 \text{ sinon.} \end{array} \right.$$

$P(Dj/Fi)$ désigne la probabilité conditionnelle de Dj sachant Fi.

Dans cette formule, un coefficient de vraisemblance positif (respectivement négatif) apparaît comme un accroissement (respectivement une diminution) de certitude pour le diagnostic Dj si le fait Fi est présent [HORVITZ 86].

Ces coefficients sont, depuis le début de leur utilisation, l'objet d'un large débat. Leurs créateurs affirment qu'ils constituent un moyen naturel pour les experts d'exprimer leur incertitude dans l'affirmation énoncée par la règle. Leurs détracteurs n'en sont pas convaincus [FIESCHI 83] et leur opposent des incohérences entre la définition probabiliste et les formules de combinaison parallèle (Tp) et séquentielle (Ts) [DEMONGEOT 88], [ROTTERDAM 86] [HECKERMAN 86]. En particulier, HECKERMAN a établi que toute interprétation probabiliste des CV nécessitait, pour être compatible avec la formule de combinaison parallèle et vérifier l'axiome de modularité (la suppression ou l'ajout de règles ne remet pas en question l'existence des autres règles), l'indépendance conditionnelle des faits Fi relativement aux Dj et à leurs complémentaires.

Un handicap important des systèmes de règles de production est de ne pas représenter explicitement les relations de dépendance et d'indépendance entre les objets du domaine. Ce problème a motivé des

recherches pour construire d'autres modèles permettant de prendre en compte l'incertitude dans les SEM. En particulier, les réseaux Bayésiens de PEARL [PEARL 88] constituent un modèle de représentation du raisonnement incertain où les dépendances et les indépendances sont clairement explicitées dans le réseau d'inférence [DEMONGEOT 88], [ROBERT 88b].

Ces coefficients de vraisemblance sont néanmoins actuellement utilisés dans de nombreux SEM (famille MYCIN, SAM, SES) avec de bons résultats comme le soulignent [BHATNAGAR 86], [HECKERMAN 86].

La mise au point des valeurs de ces CV est délicate, tant au cours de l'étape d'acquisition des connaissances, qu'à l'étape d'amélioration des performances. Modifier les CV de quelques règles pour corriger un ensemble M_f de "mal classés" pour une conclusion donnée D_f peut dégrader sensiblement les performances du Système pour les autres conclusions D_i , ($i \neq f$).

Le coefficient de vraisemblance d'une règle n'est pas sans rapport avec les observations de la base de données. La figure suivante représente un exemple de tableau ($P_k \times D$), obtenu à partir de W , en croisant une prémisse P_k avec le diagnostic D ; dans cet exemple, P_k désigne une prémisse utilisée dans une ou plusieurs règles de la base de connaissances ; D_1, D_2, D_3, D_4 sont les quatre conclusions possibles du SEM.

	D1	D2	D3	D4
Pk	40	40	20	0
VPPi ¹	0,4	0,4	0,2	0
pvi ²	0,4	0,3	0,1	0,2

¹VPPi : valeur prédictive positive de Pk pour Di, observée sur W.

²pvi : prévalence de Di observée sur W.

Règles de la base de connaissances :

SI Pk alors D1 <.3> (R1)

SI PK alors D2 <.5> (R2)

SI PK alors D3 <.2> (R3)

Ce tableau s'interprète comme suit :

a. Pk est observé 100 fois : les règles R1, R2,R3 sont donc déclenchées 100 fois.

b. $p_{v1}=VPP1$: on peut donc douter de l'utilité de la règle (R1).

c. $VPP4=0$: on peut donc proposer une règle "Pk \longrightarrow NON D4 ".

d. Si la valeur prédictive positive associée à la règle, à savoir VPPi, est supérieure à la prévalence (pvi) du diagnostic Di, on augure bien de la validité et de l'intérêt de la règle.

e. Dans notre exemple, $VPP2-pv2=VPP3-pv3$; en présence de Pk, les probabilités de D2 et D3 augmentent du même incrément. Cependant, comme $p_{v2} > p_{v3}$ avant l'information Pk, il semble logique de favoriser D2. Ceci peut être obtenu en utilisant la formule de SHORTLIFFE CV_P rappelée précédemment, à partir des données du tableau précédent.

A partir des tableaux croisés (Pk x D) (pour tout k), on peut donc déterminer :

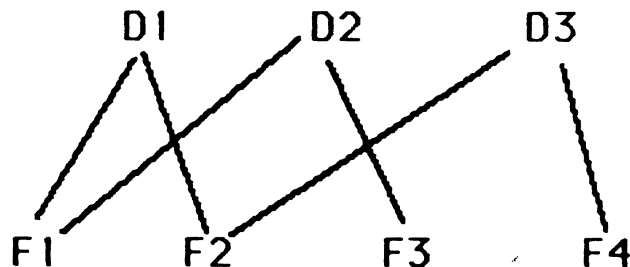
- les règles jamais ou exceptionnellement appliquées,
- les règles qui semblent inutiles, car elles ne font pas progresser le raisonnement,

- les propositions de modification de coefficients de vraisemblance par la formule CV_P,
- de nouvelles règles, absentes dans la base de connaissances. Ces règles peuvent être "positives" (affectées d'un coefficient de vraisemblance positif) ou "négatives" (elles concluent négativement à un diagnostic, et sont donc affectées d'un CVN). Ces dernières règles négatives s'avèrent généralement très utiles pour renforcer la cohérence de la base de connaissances et améliorer les performances du système expert.

7.6 Justification de la représentation à deux niveaux

Lors des étapes précédentes, nous avons considéré le problème du diagnostic comme la recherche de relations optimales entre un niveau "bas" représenté par les prémisses et un niveau "haut" représenté par les diagnostics. Cette représentation est d'ailleurs reprise par PEARL lors de l'application de réseaux Bayésiens [PEARL 87].

Bien entendu, cette approche est réductrice. Mais nous pouvons montrer que l'on peut, dans la très grande majorité des cas, se rapporter à de tels schémas à deux niveaux, au moins en ce qui concerne le raisonnement incertain :



Dans le domaine du diagnostic médical, on peut se ramener à ce schéma en appliquant l'une des opérations suivantes :

- a. Bien souvent l'étape de raisonnement incertain n'affecte qu'une des étapes de la résolution du problème. Ainsi, en Médecine, les règles qui permettent de réunir des signes en syndromes sont plus souvent des règles structurantes (un certain nombre de signes doivent être rencontrés pour évoquer le syndrome).

L'incertitude est surtout évoquée à l'étage ultérieur lorsque la combinaison de syndromes ("et/ou" éventuellement de signes) permet d'évoquer plusieurs diagnostics possibles dans un monde incertain.

On peut donc, en faisant abstraction des étapes structurantes, se ramener à un monde incertain à deux niveaux.

- b. Parfois la représentation est plus complexe et l'approche plausible ou floue affecte différents niveaux de raisonnement. Il est alors nécessaire de décomposer le savoir en plusieurs strates pour se ramener à une géométrie explorable et comparable aux éléments contenus dans la base de cas.

L'analyse d'une telle base de connaissances s'avère plus longue (car il faut multiplier les niveaux d'incertitude) et plus complexe (car la comparaison avec la base de cas est plus difficile; en effet, dans cette dernière, les syndromes du niveau intermédiaire sont vrais ou faux mais ne peuvent être considérés comme imprécis ou incertains).

Cependant, dans notre expérience, une telle décomposition est rarement nécessaire.

7.7 Confirmer les propositions de modifications par l'avis des experts

La méthodologie que nous venons de proposer a pour objectif d'améliorer les performances du système expert en termes "bien classés, mal classés". Ce critère est souvent adéquat, mais nous devons cependant en tempérer la validité par les remarques suivantes :

- l'échantillon de la base de cas W peut être biaisé, par rapport à l'utilisation réelle du système expert. Dans ce cas, une adaptation trop fine du système expert peut avoir des effets néfastes sur son utilisation en pratique courante puisque les prévalences de certaines affections (les affections rares en particulier) différeront entre W et l'espace d'application du système expert.

- En dépit de cette rareté, certains médecins veulent à tout prix faire apparaître certains diagnostics, qui, pour être rares, n'en sont pas moins d'une importance capitale (le pronostic de l'affection peut en être modifié) [GREMY 87]. Ainsi, dans le Système Expert Septicémie, les affections candidosiques ou à Listéria sont affectées d'un coefficient important qui a pour but, non de les diagnostiquer, mais d'y faire penser.

Ainsi, il est indispensable de recueillir l'avis des experts avant toute modification de la base de connaissances. Eux seuls sont en mesure de confirmer les résultats statistiques, de discuter de la représentativité de W vis-à-vis de l'espace d'application, de justifier des coefficients de vraisemblance sans rapport avec la valeur prédictive des prémisses.

8 FAUT-IL RAJOUTER DES SCORES DISCRIMINANTS ?

Dans un certain nombre de conditions, les conclusions du système expert peuvent être incomplètes. Ainsi, nous avons pu observer que seuls des sous-buts sont atteints, alors que les buts restent inaccessibles. Il semble logique d'opérer de la façon suivante :

- Regrouper un certain nombre de buts "proches" (D_{ik} , $k=1,n$) sous la forme d'un sous-but (SB_i).
- Opérer, dans un deuxième temps, une démarche de discrimination à l'intérieur du sous-but SB_i afin de diagnostiquer le but le plus probable parmi les (D_{ik}) (FIGURE 8.I).

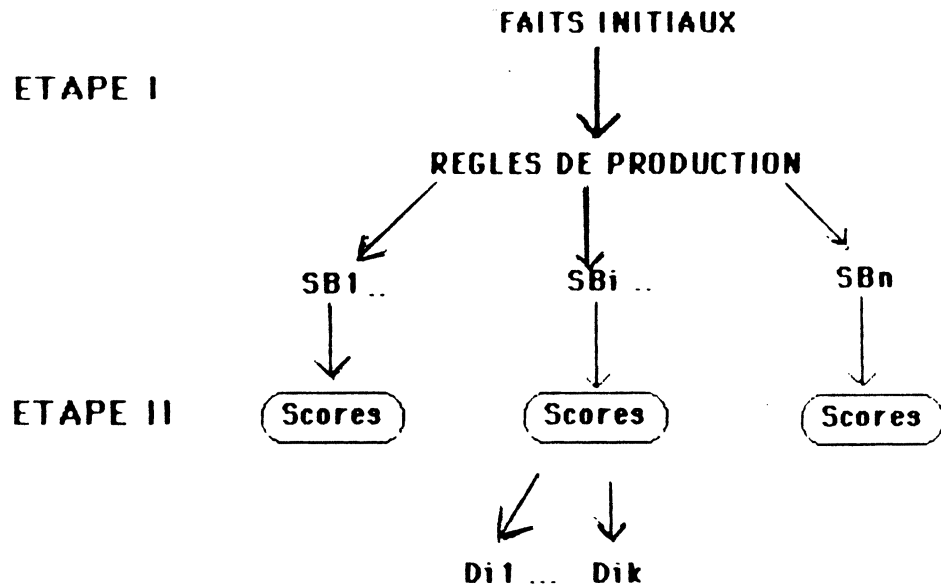


FIGURE 8.I

Cette deuxième opération devrait, dans l'idéal, être intégrée au système expert [FIESCHI 83] afin de renforcer la cohérence de la connaissance. Néanmoins, on peut aussi, dans un certain nombre de cas, faire l'hypothèse qu'une combinaison linéaire de variables

pertinentes permettrait de discriminer aisément les buts (D_i) dans une situation donnée [ROBERT 86], [ROBERT 88c]. Cette approche permettra de faire progresser facilement le raisonnement en identifiant une étape supplémentaire, permettant de passer du sous-but agrégé aux buts finaux individuels.

En s'inspirant de la méthodologie utilisée pour confirmer ou infirmer la faisabilité d'un système expert dans un domaine médical donné, on peut proposer des scores fondés sur des analyses Multifactorielles Discriminantes.

Cette approche peut paraître superflue au premier abord. En fait, l'expérience montre que les règles de production d'une part, les scores discriminants d'autre part, représentent deux formes de combinaison des mêmes connaissances et que le savoir sous une forme n'est pas toujours aisément transposable sous une autre. Leur coopération peut, dans certains cas et sous certaines conditions de validité statistique, s'avérer enrichissante.

9 EVALUER LA NOUVELLE BASE DE CONNAISSANCES SUR DES CAS NOUVEAUX

La stratégie d'amélioration de la base de connaissances a été développée sur un ensemble de cas W de la base de données. Les modifications apportées à cette base de connaissances ont amélioré ses performances sur W , en termes de pourcentage de "bien classés" et de qualité de la discrimination entre les buts D_i . La stabilité de ces améliorations devra être vérifiée en évaluant la base de connaissances sur un ensemble V de cas nouveaux ($V \cap W = \emptyset$).

10 APPLICATION DES RESEAUX BAYESIENS

Les résultats statistiques obtenus lors des étapes que nous avons décrites pourront être utilisés pour construire un réseau probabiliste dans le domaine d'application, en reprenant la méthodologie des réseaux Bayésiens [PEARL 88]. Nous disposons en effet de toute la connaissance nécessaire à la réalisation d'un tel modèle de représentation des connaissances : les liens entre les prémisses et conclusions des règles, explicités par les études statistiques, permettent de construire les noeuds et les arcs du réseau. Les données statistiques (sensibilités des signes vis-à-vis du diagnostic) fournissent les poids des arcs du réseau.

Des études multifactorielles (analyses en composantes principales) destinées à étudier les dépendances et indépendances entre les noeuds ainsi retenus pourraient apporter une aide supplémentaire pour réaliser un réseau adapté au domaine.

Il sera alors intéressant de comparer les performances de ce modèle avec celles obtenues par la base de règles initiale.

Dans notre application, nous avons montré que ces performances étaient semblables [BEUSCART 89].

11 EVALUATION SUR SITE

Toutes les étapes précédentes sont des étapes "de laboratoire". Après ces étapes "in vitro", il sera nécessaire d'affronter l'expérimentation "in vivo". Peu de systèmes experts médicaux ont, en fait, atteint un niveau de compétence et de crédibilité digne d'être proposé au grand public, qu'il soit médical ou non [FIESCHI 83],

[FERRAND 84], [BEUSCART 88a].

Dans la première phase, il sera indispensable de recueillir, après chaque interrogation, l'avis du médecin utilisateur, et d'évaluer la portée du système expert en tant qu'aide au diagnostic ou à la thérapeutique.

Selon le domaine d'application, l'évaluation de la pertinence et de l'impact du système expert sera plus ou moins facile à décrire. Ainsi, pour un système expert d'aide au diagnostic en Imagerie Médicale [BEUSCART 89], la comparaison au diagnostic réel est une chose aisée. De même, la comparaison des conclusions de SES au diagnostic bactériologique fourni 48h plus tard par le laboratoire est aussi possible.

Encore faut-il que les utilisateurs du SEM, dans le site médical où il est utilisé, acceptent de participer comme dans [FIESCHI 87] à cette étape de validation, seule capable d'affirmer l'utilité des systèmes d'aide à la décision dans le domaine d'application considéré.

CHAPITRE II : - ETUDE DE FAISABILITE D'UN SYSTEME EXPERT -

CHAPITRE II : ETUDE DE FAISABILITE D'UN SYSTEME EXPERT

1 INTRODUCTION

La mise en chantier d'un système d'aide à la décision médicale, en termes d'aide au diagnostic ou d'aide à la thérapeutique, est une tâche longue et ardue, achevée par l'étape d'évaluation et l'éventuelle validation, dont nous venons de montrer la complexité.

Aussi peut-il apparaître utile de s'assurer auparavant que le problème peut être résolu au moyen d'un raisonnement logique faisant intervenir des signes objectifs, décomposables et analysables. En effet, un programme ne pourra traiter que des informations transmissibles sous la forme de variables logiques (vrai, faux, connu, inconnu), de variables pondérées de facteurs de certitude ou de variables réelles. Toute information ne pourra être traitée par le programme qu'à la condition d'avoir été réduite à l'une de ces formes.

En médecine, certaines données sont particulièrement délicates à appréhender. C'est en particulier le fait d'items observables qui nécessitent une extériorisation par le patient et une interprétation par le médecin observateur : dépression, euphorie, angoisse, douleur, posent toujours le problème de leur expression et de l'estimation de leur intensité.

Avant de réaliser un système expert d'aide à l'interprétation des douleurs précordiales potentiellement angineuses, nous avons effectué, selon le schéma décrit en I.2, une étude de faisabilité. Comme nous allons le montrer, cette étude a permis de souligner la difficulté - voire l'impossibilité - d'une réalisation.

Une telle étude préalable nous semble être de rigueur lorsque le problème affronté est un problème ouvert qui n'a pas encore reçu, de la part de la clinique et de la paraclinique, de réponse satisfaisante et éventuellement reconnue.

La suite de ce chapitre détaille donc l'étude de faisabilité d'un système expert destiné au diagnostic de maladies angineuses face à une douleur précordiale, typique ou non.

2 LE PROBLEME MEDICAL

La maladie coronarienne (angine de poitrine, infarctus du myocarde, cardiopathie ischémique) est généralement en rapport avec une sténose artérielle coronaire. Ce rétrécissement du diamètre artériel entraîne un déséquilibre entre les apports et les besoins en oxygène du muscle cardiaque.

Lorsqu'un patient présente une lésion coronarienne, les premiers symptômes apparaissent généralement sous forme de douleur. La prévalence de lésion est en effet estimée à 4% seulement chez les patients asymptomatiques (FRAMINGHAM STUDY, [KANEL 76], [ERIKSSEN 77]). Cette douleur, communément appelée angine de poitrine, est décrite par le malade lui même : il se plaint souvent de "douleurs prenant naissance à la poitrine, déclenchées par l'effort, pouvant s'accompagner d'une sensation de mort imminente, disparaissant au repos...". Les critères de jugement du médecin sont donc essentiellement subjectifs et se fondent sur un interrogatoire du malade qui doit être minutieux, complet et non suggestif.

La coronarographie est un examen angiographique des artères coronaires, qui permet de déceler la présence ou l'absence de lésions artérielles. Grâce à cet examen, on a montré qu'un nombre important de patients se plaignant de douleurs a priori caractéristiques étaient néanmoins exempts de lésions. Plus précisément, on estime à 60% la prévalence de lésions chez les patients symptomatiques [DIAMOND 79]. Les douleurs semblables aux douleurs coronariennes peuvent avoir d'autres causes d'origine cardiaque (insuffisance ou sténose aortique, cardiomyopathie obstructive, troubles du rythme) ou non cardiaque (douleurs oesophagiennes et de la hernie hiatale, syndrome de la paroi thoracique..) [JULLIEN 83], [SAMPSON 71].

La coronarographie est cependant un examen particulièrement lourd et invasif. Elle nécessite une hospitalisation et peut comporter des risques pour le patient. Toute stratégie permettant d'aider le médecin dans son diagnostic sans y avoir recours serait donc particulièrement utile. Tout au moins serait-il souhaitable d'éviter de réaliser cet examen chez les sujets sains.

De nombreuses études ont été réalisées dans ce but. En particulier, DIAMOND [DIAMOND 79] a établi un score à partir des critères suivants :

- Douleur précordiale (+1 point).
- Douleur produite par l'effort (+1 point).
- Douleur rapidement calmée par le repos ou la trinitrine (+1 point).

Ce score permet de définir 3 classes de patients : Angor typique (3 points), atypique (2 points) et douleur non angineuse (0 ou 1 point). Des tables de prévalence de lésions selon l'âge, le sexe et le type de douleur ont ainsi pu être élaborées. Cependant, ces tables ne permettent pas toujours d'éviter la coronarographie : ainsi, le

risque de lésion s'élève encore à 30% chez les sujets âgés de plus de 50 ans souffrant de douleurs classées non angineuses.

De meilleurs résultats ont été obtenus en multipliant les tests paracliniques, tels que l'ECG d'effort ou la scintigraphie au Thallium [DIAMOND 79], [CHAITMAN 79]. Ces examens sont cependant lourds et coûteux et la plupart d'entre eux nécessite aussi une hospitalisation. Citons également des tentatives d'élaboration de scores par analyse multivariée [WEINTRAUB 85], [GREENBERG 84]. Ces scores ont été construits à partir de données générales sur le patient (âge, sexe et classification de DIAMOND) mais aussi à partir des résultats des tests précédemment cités. Leur intérêt est souvent limité dans la mesure où le pourcentage de patients correctement classés par ces scores est d'environ 70% pour des échantillons où la prévalence de lésion est de 65%.

3 VERS UN SYSTEME D'AIDE AU DIAGNOSTIC DE LA DOULEUR CORONARIENNE

Il faut remarquer qu'aucune de ces études ne prend en compte une description précise et complète de la douleur que l'interrogatoire médical permet d'obtenir : seuls les 3 critères de DIAMOND sont généralement retenus. La question à laquelle nous nous sommes efforcés de répondre était donc la suivante : une description minutieuse et complète de la douleur apporte-t-elle une information pertinente vis-à-vis du diagnostic ? Plus précisément, peut-on, à partir d'une telle description, déterminer des heuristiques, construire des scores prédictifs de douleurs, élaborer une base de connaissances à partir de ces heuristiques et de ces scores, et donc réaliser un système d'aide au diagnostic des lésions coronariennes utilisant la méthodologie des systèmes experts ? Pour ce faire, la

démarche suivante a été adoptée :

- définition par les experts des variables jugées pertinentes.
- Réalisation, sur ces variables, d'une base de cas cliniques.
- Evaluation des connaissances sur la base de cas par les techniques de l'analyse des données.
- Conclusion de l'évaluation : pertinence des données, possibilité de construire un système d'aide à la décision.

Cette démarche présente l'avantage d'obtenir, préalablement à toute tentative de construction d'un système d'aide à la décision, une évaluation de la pertinence des connaissances disponibles. On peut ainsi éviter une grande perte de temps dans la manipulation de paramètres qui, en réalité, n'apportent pas d'information suffisante. L'étude a été menée en collaboration avec l'hôpital cardiologique de LILLE (Pr LABLANCHE) [VIGNON 86].

3.1 Définition des connaissances

Comme nous l'avons relevé, la présence de lésion coronarienne est presque toujours accompagnée de l'apparition de douleurs (crise d'angine de poitrine), mais environ 40% des patients qui souffrent de douleurs présumées coronariennes sont pourtant sains. Les caractéristiques principales de cette douleur sont connues. Les experts se sont donc efforcés de décrire la sémiologie de la douleur de manière aussi minutieuse et complète que possible. En particulier, l'expertise prend en compte l'ensemble des signes généralement reconnus en faveur de la présence de lésion et ceux dont les causes ne

sont généralement pas coronariennes. 8 catégories qui correspondent à une classification sémantique de la douleur ont été définies (Annexe I) :

- a. **Siège** de la douleur. Il s'agit d'une détermination très précise des sièges possibles de la douleur. Cette douleur est généralement d'abord ressentie au thorax. Les experts ont ainsi défini une liste exhaustive des douleurs thoraciques : "milieu poitrine (premier critère de Diamond)", "sein gauche", "sein droit"... La douleur peut en plus être ressentie ailleurs qu'au thorax. Elle est alors principalement localisée aux membres supérieurs. La liste proposée permet de préciser cette localisation : "bras gauche", "poignet gauche", "coude gauche"... Cette classe comporte 14 variables non exclusives.
- b. **Irradiation** de la douleur. L'extension de la douleur est fréquente chez les patients symptomatiques [HORWITZ 74], [SAMPSON 71]. La liste définie par les experts est aussi exhaustive que possible. Elle comporte en particulier les irradiations rencontrées le plus fréquemment chez les coronariens (au bras gauche par exemple), et symétriquement, celles rencontrées le moins fréquemment (au bras droit par exemple). Les 6 variables définies ne sont pas exclusives.
- c. **Déclenchement** de la crise. Le déclenchement de la crise est un élément fondamental du diagnostic. Ainsi, une douleur produite par l'effort est le deuxième critère participant à l'élaboration du score de Diamond. Les experts ont défini une liste de 23 variables permettant de déterminer très précisément dans quel contexte la douleur apparaît le plus fréquemment : "douleur

produite par l'effort, la prise d'un repas, les rapports sexuels, douleur au repos,...". On retrouve en particulier des caractères fréquemment rencontrés chez les patients qui souffrent d'une hernie hiatale : "douleurs fréquentes la nuit, qui survient en position penchée en avant.." Ces variables ne sont pas exclusives.

- d. **Intensité de la douleur.** L'intensité de la crise est également un élément important de l'interrogatoire : la sensation d'oppression ou de mort imminente a souvent été observée chez les coronariens. Les facteurs pouvant augmenter la douleur doivent également être pris en compte. Ainsi, chez les patients qui souffrent de péricardite, la douleur augmente généralement en respirant. Cette classe comporte 8 variables non exclusives.
- e. **Type de la douleur.** Le caractère expansif, compressif ou constrictif a été fréquemment relevé chez les patients symptomatiques [SAMPSON 71]. L'appréciation de la douleur en terme de comparaison avec une douleur connue permet de prendre en compte ce caractère qui est une donnée importante de l'interrogatoire. Ainsi, un douleur "comparable à un coup de poignard" est fréquemment rencontrée chez les patients qui présentent un syndrome radiculaire. La liste comporte en outre 8 items tels que "douleur comparable à un étau, à un corset, à un barre...". Ces items ne sont pas exclusifs.
- f. **Durée de la crise et cause de son interruption.** La durée de la crise et son interruption sont des éléments importants du diagnostic. Une douleur très brève ou au contraire qui se prolonge des journées entières se rencontre fréquemment chez les patients non coronariens. Les experts ont défini 8 classes

exclusives permettant de déterminer précisément la durée de la crise. La disparition rapide de la douleur par la cessation de l'effort ou la prise de trinitrine est le troisième critère du score de Diamond. Outre les facteurs pouvant provoquer la cessation de la crise (arrêt de l'effort par exemple), 5 classes exclusives permettant d'apprécier le temps de cessation ont été définies par les experts.

- g. **Autres signes cliniques accompagnant** éventuellement la douleur. Il s'agit de signes témoignant de l'impact émotionnel de la douleur (sueurs, palpitations) ou de signes rencontrés chez les coronariens et qui peuvent égarer le diagnostic : vomissements, nausées. 5 signes non exclusifs ont été retenus par les experts.
- h. **ancienneté des douleurs, nombre d'hospitalisations, fréquence.** Les experts ont jugé utile de prendre en compte l'histoire de la douleur : l'ancienneté et le nombre d'hospitalisations ont été évalués par 4 classes. Enfin, 6 classes exclusives permettent de déterminer la fréquence actuelle des crises.

Les experts ont ainsi défini 120 variables binaires (réponse oui/non) donnant une description aussi précise et complète que possible de la douleur. Ces variables ont été regroupées pour constituer un questionnaire destiné à être rempli par les malades (**Annexe I**). Pour minimiser les erreurs d'interprétation et les problèmes de compréhension, les questions ont été formulées de manière simple et précise. Le questionnaire a été structuré en 21 questions principales clairement mises en évidence. Pour chacune de ces questions, le patient pouvait choisir une ou généralement plusieurs réponses dans la liste proposée. Il s'agissait donc d'examiner la

possibilité de définir, à partir de ce questionnaire, des profils de douleur coronarienne et non coronarienne.

3.2 La base de cas

179 patients ont été recrutés prospectivement à l'hôpital cardiologique de LILLE entre les 1/10/84 et 30/5/85 pour participer à cette étude. Ces malades souffraient de douleurs présumées coronariennes et avaient été adressés en cardiologie en vue d'un bilan coronarographique. Les patients présentant des antécédents connus d'infarctus ou une valvulopathie avait préalablement été exclus.

Le questionnaire a été distribué aux patients pour être rempli par eux-mêmes avant la réalisation de la coronarographie, l'intervention du médecin se limitant à d'éventuelles explications sur les questions posées. Cette démarche a permis de collecter des réponses caractérisant réellement la douleur ressentie par le malade sans qu'elles soient filtrées par le médecin.

Ces données, complétées d'informations cliniques (âge, sexe..) et du résultat de la coronarographie, ont été archivées anonymement sur le VAX 780 du CERIM pour être analysées. Notre échantillon comportait 145 hommes et 34 femmes. La prévalence de la lésion et surtout les caractéristiques de la douleur étant différentes selon le sexe, ces 2 populations devaient être analysées séparément. Le nombre de femmes recrutées étant insuffisant pour être représentatif, l'étude a été limitée à la population masculine.

4 L'ETUDE DE FAISABILITE

4.1 Par analyse univariée

Une analyse univariée a tout d'abord permis de vérifier la représentativité de notre population :

- Age = 51,2 +/- 9,2 ans.
- Prévalence de lésions = 64%.

Le lien entre les variables du questionnaire et le diagnostic a ensuite été étudié par analyse bivariée. A chaque variable V du questionnaire, nous avons associé le tableau de fréquences VxD, D désignant le résultat de la coronarographie. Nous avons ainsi calculé les paramètres classiques : sensibilité, spécificité et valeurs prédictives. Le TABLEAU 4.I montre les meilleurs résultats obtenus par cette analyse bivariée.

DOULEURS	VPP %	VPN %	REPOSES POSITIVES %	CHIDEUX
PRODUITE PAR RAPPORTS SEXUELS	89	40	13	NS
COMPARABLE PINCEMENT	79	38	13	NS
COMPARABLE BRULURE	79	41	28	NS
AU DEBUT EFFORT	78	38	16	NS
PRES SEIN DROIT	77	37	9	NS
PRES COUDE DROIT	76	37	12	NS
PRODUITE PAR EXPOSITION FROID	76	57	23	NS
PRODUITE PAR MARCHE	73	43	43	P<5%
COMPARABLE CRAMPE	73	37	15	NS
PRODUITE MATIN-LEVER	67	58	28	NS
ENTRAINE OPPRESSION	55	21	61	P<1%

TABLEAU 4.I

Seules 2 variables caractéristiques de la douleur sont significativement liées au diagnostic : l'absence de sensation d'oppression quand la douleur se manifeste et le déclenchement de la douleur par la marche semblent favoriser la présence de lésion (au risque p , respectivement: $p < 1\%$ et $p < 5\%$). Pour les autres variables ne figurant pas dans ce tableau, les valeurs prédictives positives (respectivement négatives) sont voisines de la prévalence p (respectivement $1-p$). Prises individuellement, les variables sont donc peu prédictives.

Naturellement, cette étude ne permet pas de conclure. Ce sont en effet très souvent des associations de symptômes qui permettent de prédire le diagnostic. Des études multifactorielles nous ont permis d'étudier si de telles associations pertinentes pouvaient être mises en évidence.

4.2 Evaluation des connaissances par analyses multifactorielles

Avec l'accord des experts, les variables redondantes et insuffisamment documentées ont été éliminées. Les analyses ont ainsi été effectuées sur 58 symptômes et la variable lésion (présence ou absence) (TABLEAU 4.II).

DOULEURS	DOULEURS
MILIEU POITRINE	PRODUITE SOIR COUCHER
SEIN GAUCHE	PRODUITE POSITION AVANT
SEIN DROIT	ENTRAINE GENE
DEUX SEINS	ENTRAINE OPPRESSION
DANS LE DOS	ENTRAINE ANGOISSE
AILLEURS THORAX	ENTRAINE SENSATION MORT
BRAS GAUCHE	COMPARABLE CRAMPE
BRAS DROIT	COMPARABLE COUP DE POING
EPAULE GAUCHE	COMPARABLE PIQURES AIGUILLES
EPAULE DROITE	COMPARABLE ETAU
COUDE GAUCHE	COMPARABLE CORSET
COUDE DROIT	COMPARABLE BRULURE
POIGNET GAUCHE	COMPARABLE COUP DE GRIFFE
POIGNET DROIT	COMPARABLE PINCEMENT
ETENDUE COU	COMPARABLE COUP DE BARRE
ETENDUE JOUES	DUREE > 10 MINUTES
ETENDUE MACHOIRES	PRISE TRINITRINE
ETENDUE BRAS GAUCHE	TEMPS CESSATION < 10 MINUTES
ETENDUE BRAS DROIT	ACCOMPAGNEE ACIDE BOUCHE
ETENDUE ESTOMAC	ACCOMPAGNEE SUEURS
PRODUITE MARCHÉ	ACCOMPAGNEE SENSATION COEUR RAPIDE
PRODUITE MONTEE ESCALIER	ACCOMPAGNEE PERTE DE CONNAISSANCE
PRODUITE RAPPORTS SEXUELS	SURVIENT A L'EFFORT
PRODUITE EMOTION	SURVIENT AU REPOS
PRODUITE REPAS	SURVIENT DEBUT EFFORT
PRODUITE MARCHÉ APRES REPAS	PLUS FREQUENTE AU REPOS
PRODUITE EXPOSITION FROID	PLUS FREQUENTE A L'EFFORT
PRODUITE EXPOSITION VENT	PLUS FREQUENTE LA NUIT
PRODUITE MATIN LEVER	ACCOMPAGNEE DE NAUSEES

TABLEAU 4.II

4.2.1 Analyse globale

Les 58 variables et les 145 individus constituent une matrice 145x58. Les variables étant binaires, le nuage des individus appartient à un hypercube de R^{58} . Notre objectif est tout d'abord exploratoire : mesurer la dispersion du nuage, observer les positions relatives entre les individus sains et pathologiques, découvrir d'éventuels sous-groupes d'individus semblables, étudier les corrélations entre les symptômes. L'Analyse en Composantes

Principales (ACP) est une des techniques de la statistique descriptive multidimensionnelle particulièrement adaptée à la représentation synthétique des grands tableaux de données [BENZECRI 81]. L'interprétation des différents graphiques fournis par cette analyse permet de répondre à notre objectif.

Une ACP canonique (non normée) sur les 58 variables et 145 sujets a tout d'abord été réalisée. La variable lésion a été projetée en variable supplémentaire.

La première valeur propre explique 13,6% de l'inertie totale du nuage. Ce résultat assez faible doit cependant être relativisé dans la mesure où une variable n'explique à elle seule que 1,7% (1/58) de cette inertie. Le pourcentage de reconstruction du nuage dans le premier plan principal est de 21,5%. Le nuage des individus n'est donc pas réparti de façon homogène dans toutes les directions de R^{58} . Les valeurs propres suivantes sont cependant peu explicatives (FIGURE 4.I).

SOMME DES VALEURS-PROPRES

9.74259472

HISTOGRAMME DES PREMIERES VALEURS-PROPRES

	VALEUR-PROPRE	POURCENTAGE	POURCENTAGE CUMULE	
1	1.32329483	13.58	13.58	*****
2	0.75257670	7.33	21.41	*****
3	0.54840416	5.63	27.04	*****
4	0.49477350	5.08	32.12	*****
5	0.41600472	4.27	36.39	*****
6	0.38893220	3.99	40.38	*****
7	0.36545697	3.75	44.13	*****
8	0.31330401	3.22	47.35	*****
9	0.30186445	3.10	50.44	*****
10	0.29152557	2.99	53.44	*****
11	0.26454052	2.72	56.15	*****
12	0.23783913	2.44	58.59	*****
13	0.22717035	2.33	60.93	*****
14	0.21728154	2.23	63.16	*****
15	0.20674907	2.12	65.23	*****

Figure 4.I ACP (58 variables, 145 individus)

La FIGURE 4.II montre la projection des individus dans le premier plan principal. On remarque une très grande dispersion dans toutes les régions du plan, entre les individus sains et pathologiques. Il en est de même dans les plans principaux suivants (1,3), (1,4) et (1,5).

La représentation des variables dans le plan des 2 premières composantes principales donne peu de renseignements. Les variables sont très dispersées ou mal représentées dans ce plan. Cependant, pour les variables bien représentées, quelques proximités naturelles sont mises en évidence :

- douleurs d'effort et produites par la marche ($r=0,52$), douleurs d'effort et produites par la montée d'un escalier ($r=0,59$).
- douleurs bras gauche (respectivement bras droit) et étendues bras gauche (respectivement bras droit), $r=0,83$ (resp. $r=0,88$).

Ces corrélations sont naturelles et attendues, elles témoignent d'un remplissage correct du questionnaire par les patients.

- on remarque par ailleurs qu'une douleur au côté droit est généralement diffuse : les variables douleurs au bras droit, épaule droite, coude droit et poignet droit sont proches. Cette propriété ne se retrouve pas pour les douleurs au côté gauche.

Aucune variable n'est fortement corrélée avec les différents axes. La variable lésion projetée en variable supplémentaire est spectaculairement mal représentée dans tous les plans successifs.

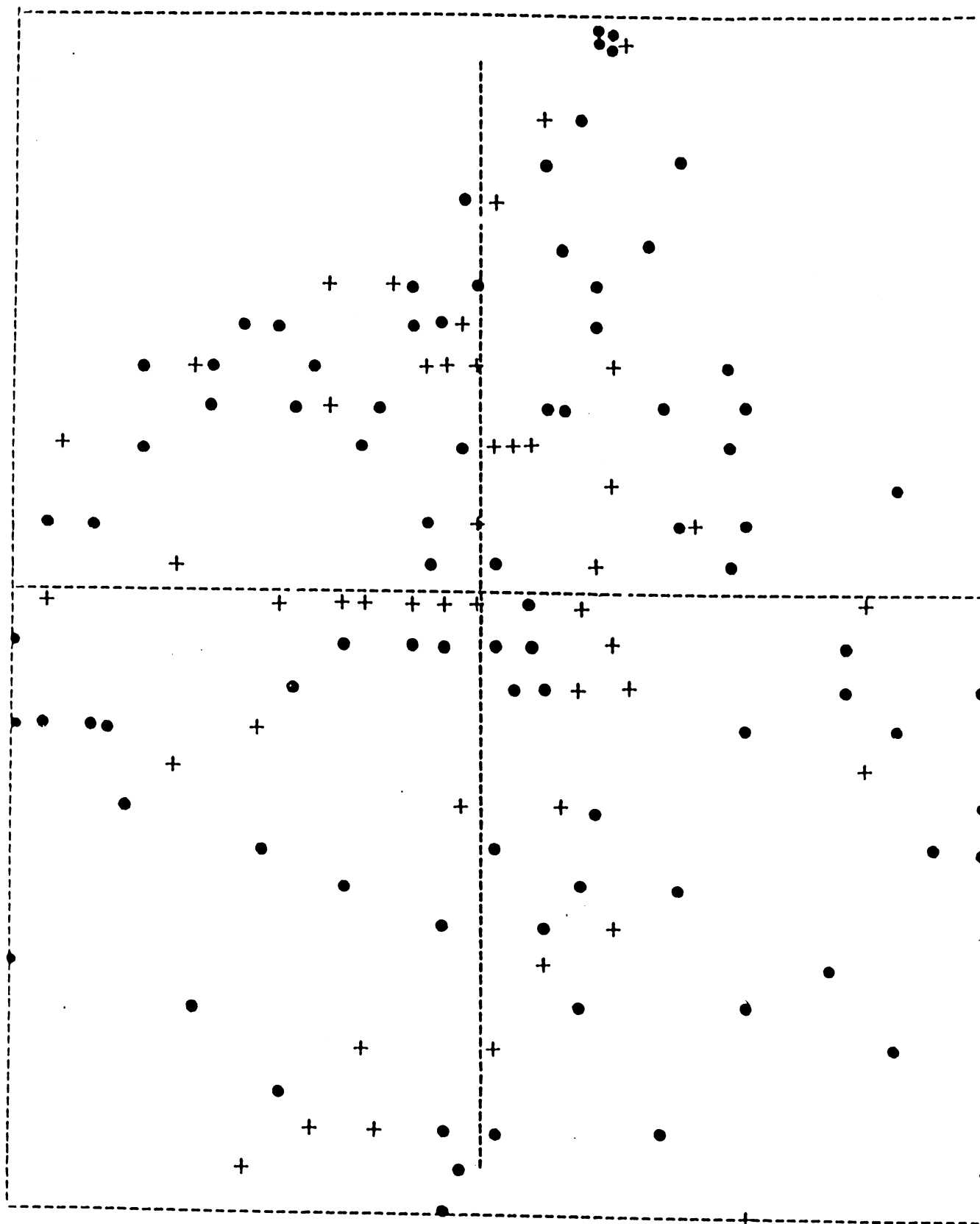


Figure 4.II (ACP 58 variables, 145 individus)

● . "Πόλιον" + . "ΜΝΟΝ Ιόλκίον"

4.2.2 Etude des groupes sémantiques.

Nous avons ensuite réalisé des ACP sur des sous-espaces de R⁵⁸ correspondant à la classification sémantique de la douleur donnée par les experts. 7 classes ont été analysées :

1. douleurs au thorax.
2. douleurs hors du thorax.
3. extension de la douleur.
4. type de douleurs.
5. intensité de la douleur.
6. déclenchement de la douleur.
7. signes d'accompagnement.

La FIGURE 4.III montre la projection des individus dans le premier plan principal (60% de reconstruction de l'inertie totale) pour l'analyse du groupe des douleurs au thorax. On observe des points isolés et des groupes d'individus confondus. Après contrôle sur le fichier des données, il s'agit ici d'individus réellement confondus. Il apparaît ainsi 3 profils de douleur au thorax : douleur uniquement au milieu de la poitrine (30% des cas), milieu de la poitrine et 2 seins (20%), milieu de la poitrine et sein gauche (12% des cas). Ces profils sont visiblement sans rapport avec le diagnostic de lésion : chaque groupe comporte environ 50% de patients pathologiques. On remarque une très grande dispersion dans toutes les régions du plan entre les individus sains et pathologiques.

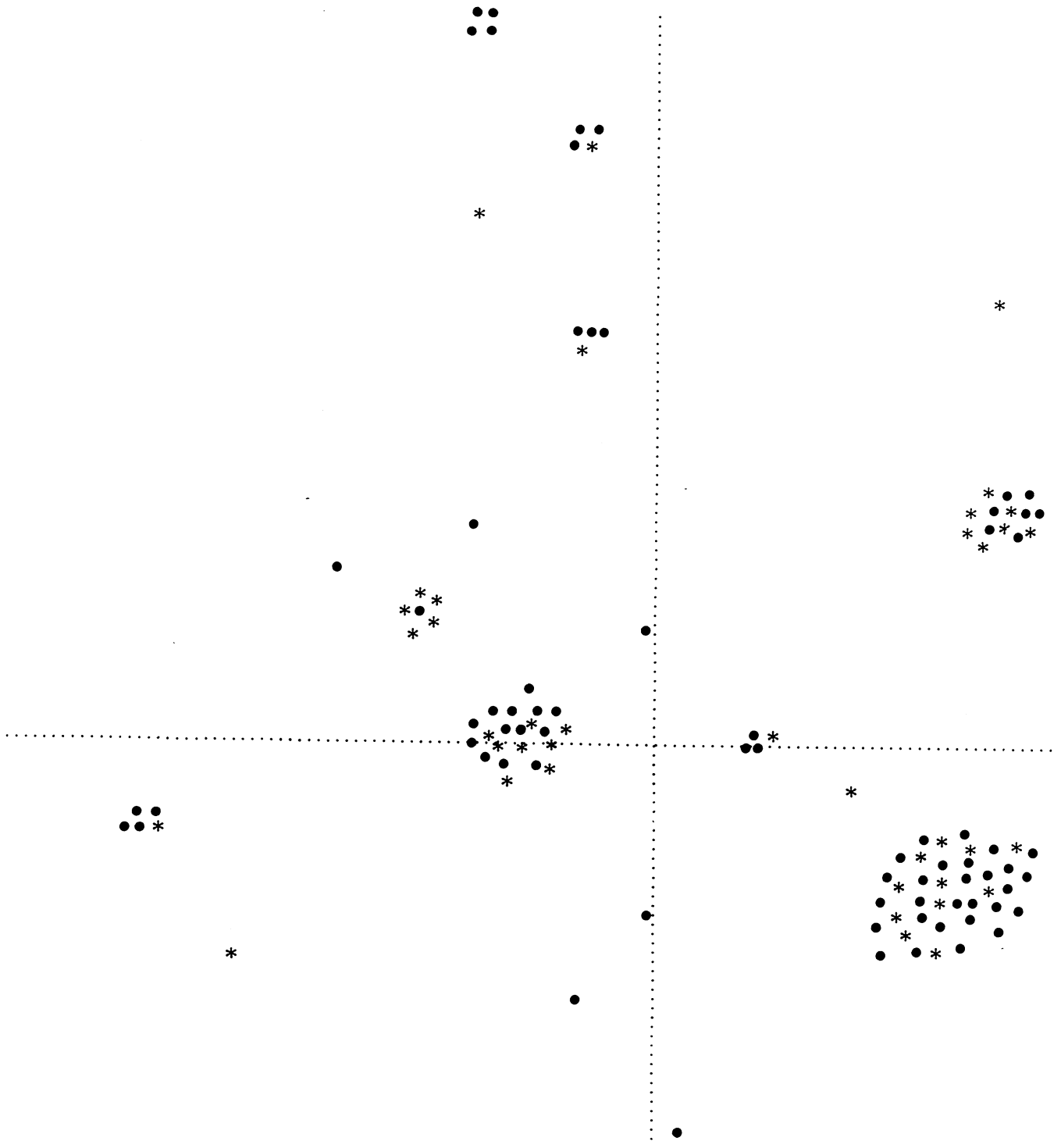
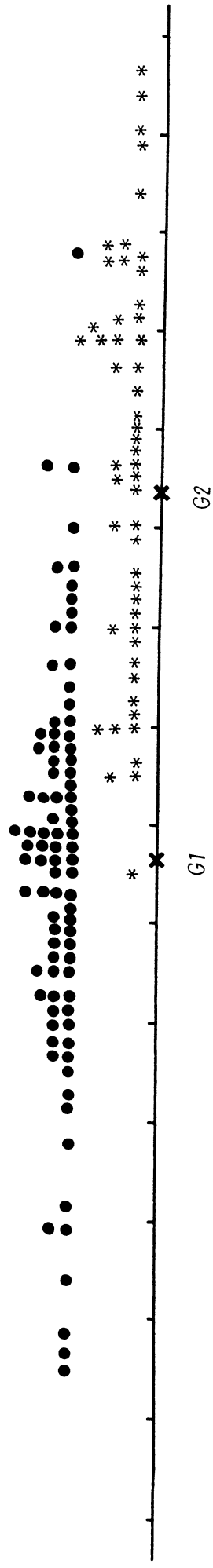


Figure 4.III ACP (Douleurs Thorax)
 ● "Lésion" * "Non Lésion"

Pour les autres ACP, les résultats peuvent être résumés de la manière suivante :

- 4 profils ont également été mis en évidence par l'analyse du groupe "type de douleur" : douleurs comparables uniquement à une brûlure, à une barre ou à un étau, et douleurs comparables à la fois à un étau et à une barre.
- pour les douleurs hors thorax, il existe un groupe de patients pour lesquels la douleur ne s'irradie que vers le bras gauche (13% des cas).
- nous n'avons pas retrouvé de tels profils pour les autres classes.
- les différentes projections permettent de plus de constater l'existence de groupes d'individus n'ayant répondu à aucune des questions proposées (individus asymptomatiques pour la classe étudiée) : 47% de non réponse pour la classe "hors thorax", 42% pour l'extension des douleurs et signes d'accompagnement, 21% pour les facteurs déclenchants.

Comme pour la classe "douleurs au thorax", les groupes d'individus mis en évidence, profils et patients asymptomatiques sont sans rapport avec le diagnostic. Les individus ayant le même profil sont différents pour chaque classe : nous n'avons pas mis en évidence de profils globaux. Pour les points isolés, on observe également une dispersion totale entre les individus sains et pathologiques.



● "Lésion (Centre de Gravité G1) * "Non Lésion" (Centre de Gravité G2)

4.2.3 Analyse discriminante.

Les profils de douleurs et les composantes principales ne permettant pas de distinguer les individus sains des individus pathologiques, nous avons tout d'abord réalisé une analyse factorielle discriminante [ROMEDER 73] [NAKACHE 80] sur l'ensemble des variables et des individus. Nous cherchons à différencier 2 groupes : "lésion" et "non lésion". L'unique axe factoriel discriminant correspond donc à la fonction discriminante de FISHER. Cet axe discriminant est la combinaison des 58 variables participant à l'analyse, dont les valeurs sont les plus différentes possibles pour les individus sains et pathologiques.

La FIGURE 4.IV montre la répartition des individus sur l'axe factoriel discriminant. Cet axe passe par les centres de gravité des deux groupes "lésion" et "non lésion". Ce graphique permet de visualiser la qualité de la séparation de ces deux groupes par la fonction discriminante. Elle sera d'autant meilleure que les centres de gravité seront éloignés et que les individus de chaque groupe seront proches de leur centre de gravité. On peut alors faire les observations suivantes :

- a. Les centres de gravité sont proches.
- b. Les groupes "lésion" et "non lésion" sont dispersés autour de leurs centres de gravité.
- c. Les deux groupes se superposent dans la zone centrale du graphique : cette zone correspond à environ 20% des patients.

Nous avons étudié les corrélations entre les variables initiales et l'axe discriminant afin de mettre en évidence celles qui interviennent le plus pour séparer les deux groupes "lésion" et "non lésion". On constate alors qu'aucune variable n'est fortement liée à cet axe. Les types de douleurs ayant les plus fortes corrélations positives sont D1 "douleur entraînant une oppression" ($r=0,34$), D2 "douleur entraînant une angoisse" ($r=0,26$) et D3 "douleur accompagnée de remontée de liquide acide dans la bouche" ($r=0,22$). Ceux qui ont les plus fortes corrélations négatives sont D4 "douleur produite par la marche" ($r=-0,23$), D5 "douleur produite par les rapports sexuels" ($r=-0,29$) et D6 "douleur comparable à une brûlure" ($r=-0,26$). Ces différentes corrélations étant peu élevées, il est difficile d'interpréter la fonction discriminante. Il semble que l'association des trois signes D1, D2, et D3, se retrouve en moyenne, plus souvent chez les sujets sains, alors que l'association des signes D4, D5 et D6, s'observe plus souvent chez les sujets atteints d'une lésion coronarienne.

Nous avons ensuite réalisé une analyse discriminante pas à pas afin d'extraire le sous-ensemble des variables assurant la meilleure discrimination entre les 2 groupes. Le TABLEAU 4.III donne la liste de ces variables, les coefficients de la fonction discriminante et le classement obtenu. Le pourcentage de biens classés est de 70%, la valeur prédictive positive de 72% et la valeur prédictive négative de 64%.

DOULEURS Xi	COEFFICIENTS Ki
COMPARABLE PINCEMENT	+ 2,5
PRODUITE RAPPORTS SEXUELS	+ 2
COMPARABLE BARRE	+ 1,1
COMPARABLE BRULURE	+ 1,1
COMPARABLE CRAMPE	+ 1,1
PRISE TRINITRINE	+ 1
ENTRAINE GENE	- 1
PRODUITE EXPOSITION VENT	- 1,2
ENTRAINE ANGOISSE	- 1,1
ENTRAINE OPPRESSION	- 1,5
ACCOMPAGNEE ACIDE BOUCHE	- 1,2
TEMPS CESSATION < 10 MINUTES	- 2
COMPARABLE COUP DE GRIFFE	- 2,4
COMPARABLE COUP DE POING	- 2,1
CONSTANTE (C)	+ 3,1

Score S= C+ KiXi;

	LESION	NON LESION
S > 0	82	32
S < 0	11	20
	93	52

TABLEAU 4.III

5 CONCLUSION DE L'ETUDE DE FAISABILITE.

L'analyse en composantes principales est une première approche pour explorer les données et juger de leur pertinence vis-à-vis du but recherché.

Nous avons réalisé ici une ACP non normée. En effet, il est nécessaire de normer les variables lorsqu'elles sont hétérogènes, de façon à ne pas privilégier celles dont la variance élevée s'explique avant tout par un phénomène d'échelle (par ex. âge et température). S'agissant de variables binaires, la variance est toujours inférieure

à 1 (en fait 0,25); une normalisation privilégierait donc les variables de faible variance [FOUCARD 85].

Il faut distinguer pour l'analyse les variables explicatives (symptômes) de la variable à expliquer (lésion). Cette dernière ne doit pas être prise en compte dans le calcul des composantes principales. Nous l'avons projetée en variable supplémentaire pour étudier les liaisons diagnostic/symptômes et diagnostic/composantes principales.

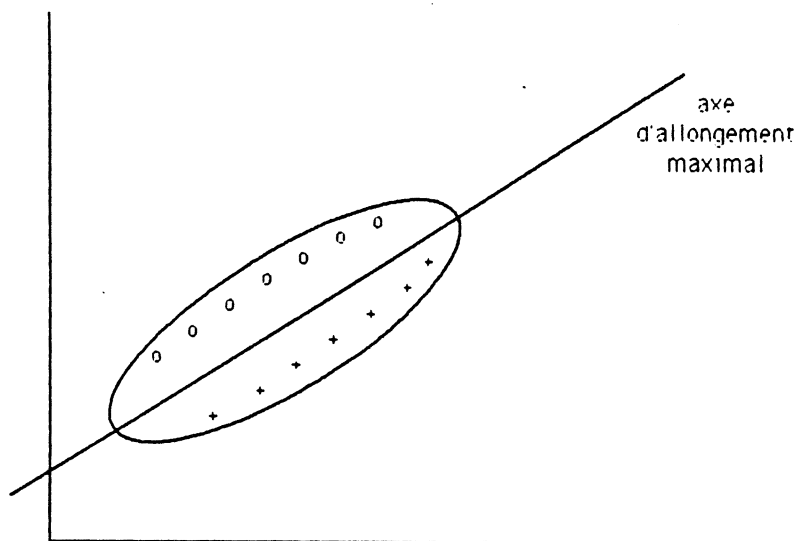
Le premier plan principal reconstruit 21,5% de l'inertie totale du nuage (3,5% (2/58) en moyenne pour 2 variables). Les 2 premières composantes principales représentent donc des directions privilégiées pour la dispersion. Ce pourcentage de reconstruction est cependant insuffisant pour se limiter à une étude dans le premier plan principal et les projections dans les plans (1,1) à (1,5) ont également été examinées. Les valeurs propres suivantes ne sont plus significatives.

L'étude des contributions des variables à la formation des composantes principales ne fait pas apparaître de sous-groupe de variables responsable de la formation de ces axes. Nous n'avons donc pas mis en évidence d'association de symptômes expliquant la dispersion des individus dans les directions où cette dispersion est maximale.

D'autre part, la variable lésion est très mal représentée dans les plans successifs, ce qui signifie que les composantes principales sont sans rapport avec le diagnostic.

Enfin, les projections des individus dans les plans principaux examinés montrent une totale dispersion dans toutes les régions du plan entre les individus sains et pathologiques.

Cependant, compte tenu du faible pourcentage de reconstruction du nuage, une telle réduction de R^2 à R^2 peut cacher bien des choses et il était nécessaire de poursuivre les analyses afin d'éviter toute conclusion hâtive. En particulier, il peut exister une combinaison linéaire des variables permettant de déduire le diagnostic et qui soit sans rapport avec les premiers axes d'allongement maximal du nuage comme le montre la figure suivante:



Nous avons donc étudié le nuage des individus dans des sous-espaces de R^{58} . Ces sous-espaces correspondent aux différents types de douleurs décrits par les experts. Ces analyses ont permis de mettre en évidence des profils de douleur et des groupes d'individus asymptomatiques: ils n'auraient pas pu être identifiés à partir du tableau des données (matrice 149x58).

Les différentes projections ont montré que ces groupes et profils étaient sans rapport avec la présence ou l'absence de lésion (FIGURE 4.III). Dans chacune des classes étudiées, nous n'avons donc pas mis en évidence des associations de symptômes liées au diagnostic.

L'étape descriptive réalisée par les différentes ACP ne nous a pas permis de déterminer des sous-groupes de variables pertinentes pour le diagnostic de lésion. Pour examiner la qualité de la séparation des 2 groupes par des combinaisons linéaires des variables symptôme, nous avons réalisé une analyse factorielle discriminante. On cherche dans cette analyse l'hyperplan permettant de différencier le mieux possible les 2 groupes.

L'axe discriminant (FIGURE 4.IV) semble séparer les 2 groupes. Cependant, on remarque que les individus de chaque groupe sont dispersés autour de leur centre de gravité. D'autre part, ces centres de gravité sont proches et l'on observe une zone, proche des centres de gravité, où les individus sains et pathologiques se superposent (20% des patients). L'étude des corrélations entre les variables et l'axe discriminant montre qu'aucune variable n'est fortement corrélée à cet axe. On ne peut donc pas mettre en évidence de sous-groupes de variables responsables de l'axe expliquant la partition individus sains et pathologiques.

Nous savons que la fiabilité de la discrimination est meilleure si l'on considère un petit nombre de variables : il est plus facile de trouver un hyperplan séparateur dans un espace de dimension élevé (ici l'ensemble des variables). Nous avons donc réalisé une analyse discriminante pas à pas (TABLEAU 4.III). Un sous-groupe de 14 variables les plus explicatives a été mis en évidence. Cependant, les résultats

du classement par la fonction discriminante amènent les remarques suivantes :

- le pourcentage de bien classés est de 70%. Une simple affectation au hasard tenant compte des prévalences de 2 groupes conduirait à un pourcentage de bien classés de 54%.
- la valeur prédictive positive est de 72%, elle est donc proche de la prévalence de l'échantillon. Considérons le critère de décision défini par la fonction discriminante, D^1 : "lésion si $S > 0$ ", et D^2 "non lésion si $S < 0$ " (TABLEAU 4.III). L'erreur commise en déclarant " $S > 0$ entraîne lésion" est égale au pourcentage de faux positifs. En assimilant l'absence de lésion à l'hypothèse nulle de la théorie des tests [LAPLANCHE 87], cette erreur correspond à l'erreur de première espèce et vaut 61%. En d'autres termes, un nombre très important de patients sains subirait la coronarographie sur la base de ce score, avec les conséquences que nous avons relevées au début de ce chapitre : examen invasif, coût élevé,...

La fonction discriminante ne permet donc pas de différencier de manière suffisamment significative les 2 groupes.

6 CONCLUSION GENERALE

Ces résultats ne permettent pas de dégager des items ni des combinaisons linéaires d'items pertinents pour le diagnostic de maladie coronarienne. Il n'est donc pas possible de mettre en relation des signes ou associations de signes avec un diagnostic possible de coronarite.

Il semble difficile de proposer, dans ces conditions, des règles d'inférence pertinentes, puisque la probabilité de la coronarite n'est guère différente selon les syndromes douloureux cliniques.

Ce résultat n'était pas prévisible. Nous aurions au moins aimé confirmer la validité du score de DIAMOND. Bien que valide sur la côte EST des USA, ce score n'apparaît pas pertinent pour les malades du Nord de la France.

Pourtant, l'intérêt de ce système expert en terme de santé publique était évident : en raison de la fréquence des maladies coronariennes et de la variabilité des tableaux cliniques, il eût été intéressant de disposer d'un outil d'aide au diagnostic, par essence non invasif, qui puisse orienter la stratégie d'exploration à partir des caractères cliniques.

Cette expérience a démontré l'intérêt, dans ces conditions, des études de faisabilité pour évaluer la pertinence des connaissances par rapport au diagnostic recherché.

CHAPITRE III A : - PRESENTATION DU SYSTEME EXPERT SEPTICEMIE -

CHAPITRE III A : PRESENTATION DU SYSTEME EXPERT SEPTICEMIE

1 INTRODUCTION

Lorsqu'un système expert est considéré comme faisable, sa réalisation nécessite :

- le choix d'une représentation des connaissances,
- la détermination des items pertinents,
- le choix de l'expression des conclusions.

Mais il est nécessaire également de prévoir, à ce stade, les méthodes par lesquelles on pourra s'assurer de la validité des conclusions fournies par le système expert. Pour cela, il semble utile de concevoir simultanément une base de cas qui permettra la simulation de problèmes réels, tels qu'ils sont posés en pratique. L'articulation entre items de la base de cas et la base de connaissances favorisera cette simulation.

De la qualité clinique et épidémiologique de la base de cas dépendra aussi la validité de l'évaluation du système expert. Plus le champ de recueil sera large, plus les problèmes soumis seront divers et meilleure sera l'évaluation.

Nous avons mis en oeuvre cette démarche lors de la réalisation du Système Expert Septicémie et au cours de son évaluation.

Nous détaillerons donc ci-après :

- la nature du Système Expert Septicémie,
- son évaluation,
- les étapes d'amélioration et de modification qui ont été

entreprises, ainsi que leur impact sur les performances du système.

2 LE PROBLEME MEDICAL

La définition de la septicémie est toujours celle donnée par GASTINEL ET REILLY en 1927 :

"L'état septicémique correspond à toute infection générale conditionnée par des décharges massives et répétées dans le sang de bactéries pathogènes et de leur poison. Issue d'un foyer septique, appréciable ou non, cette migration de germes, continue ou discontinue, engendre des signes généraux graves tenant à de multiples embolies microbiennes, à l'action des toxines bactériennes, enfin aux effets nocifs des produits de désintégration cellulaire, tous symptômes laissant au second plan le foyer infectieux initial".

Pour un exposé simple et concis sur la septicémie, on pourra consulter [MODAI 84].

Les paramètres de la septicémie sont relativement bien connus : antécédents du patient, localisations septiques (primaires et secondaires), lieu d'acquisition, température... Ces données peuvent donc être décrites et enregistrées dans le cadre d'un dossier médical informatisé.

La septicémie nécessite une hospitalisation. C'est une maladie infectieuse grave face à laquelle le pronostic vital est souvent en jeu. Son évolution favorable est liée à l'institution, le plus rapidement possible, d'un traitement efficace contre le germe responsable de l'infection et adapté au patient.

Un tel traitement ne peut être institué qu'après avoir résolu les deux problèmes complexes suivants :

- **Identification du germe** : les résultats bactériologiques permettant d'identifier le germe pathogène ne sont souvent disponibles qu'après 48 ou 72 heures, période au cours de laquelle des complications graves (dissémination septique, choc), voire le décès peuvent survenir. Ce sont alors les seuls signes cliniques et biologiques qui peuvent guider le diagnostic du praticien et lui permettre de déterminer le germe le plus probablement responsable de l'infection.

- **Antibiothérapie adaptée** : la connaissance du germe pathogène ne résoud pas tous les problèmes. En effet, si le choix du traitement antibiotique est, bien sûr, fonction de ce germe, il dépend également du tableau clinique présenté par le patient : pour un même agent infectieux, le choix des antibiotiques, leur posologie et leurs associations possibles varient selon les localisations septiques et le terrain général du patient.

Dans de nombreux cas, les médecins non spécialistes (de ce domaine) n'ont pas l'expérience en infectiologie et antibiothérapie qui leur permettrait de résoudre ces problèmes. Ce sont essentiellement des cliniciens hospitaliers spécialistes en maladies infectieuses qui détiennent les connaissances requises.

Les connaissances mises en jeu par ces experts sont vastes et incertaines. Ce sont des connaissances épidémiologiques et heuristiques, caractérisées par une grande évolutivité [MCGOWAN 75], [BABEAU 85] : par exemple, les maladies induisant une

immuno-dépression (sida, chimiothérapie anticancéreuse) créent un terrain favorable aux septicémies [MODAI 84]. Les modes de raisonnement que ces experts utilisent sont cependant analysables et reproductibles.

Ces différentes remarques justifient la création d'un système d'aide à la décision en cas de septicémie ou de maladie infectieuse grave utilisant la méthodologie des systèmes experts et ont conduit à la mise en place du PROJET SES (Système Expert Septicémie) [BEUSCART 87b, 88a].

Le PROJET SES a été réalisé en tenant compte en particulier de deux facteurs :

- a. **Obtenir une expertise la plus large possible.** Dans certains domaines de la médecine et en particulier dans le domaine de l'infectiologie, le savoir n'est pas concentré au niveau d'un expert ou de quelques experts d'un service donné : face à une même pathologie, les avis des spécialistes peuvent diverger. Dans de tels domaines, un système expert ne sera reconnu et utilisé que si sa réalisation a été le fruit d'un consensus de la communauté des experts. De plus, dans le domaine de l'infectiologie, une expertise limitée à un service pourrait induire d'importants biais ayant pour conséquence une réduction de la validité du système à une région donnée : par exemple, la brucellose est plus fréquente dans le Sud-Ouest de la France, tandis que les septicémies consécutives à des pneumopathies graves sont plus fréquentes dans le Nord.

- b. **Prévoir une utilisation décentralisée du système.** Il est nécessaire de prendre en compte l'aspect éparpillé de la communauté médicale : de nombreux systèmes experts sont peu utilisés en raison de la difficulté qu'ont les utilisateurs potentiels à y accéder [REGGIA 85].

Le PROJET SES est né en 1985, grâce à la collaboration de 19 centres français hospitalo-universitaires et du CERIM, sous l'égide de la Ligue pour la Prévention des Maladies Infectieuses (FIGURE 2.I). Un groupe de travail, le **GROUPE SES**, s'est alors constitué, associant 40 experts issus de ces centres universitaires de Maladies Infectieuses et le CERIM. Le groupe SES s'est fixé les objectifs suivants :

- a. La réalisation d'un dossier médical informatisé et décentralisé destiné à améliorer la connaissance des maladies infectieuses et de leur évolution et constituant une base de cas pour valider le système expert : La Base de Données SES (**BD_SES**).
- b. La réalisation et la validation, à partir d'une **expertise multiple**, d'un système expert d'aide au diagnostic et à l'antibiothérapie, accessible par le réseau télématique [BEUSCART 88b] : Le Système Expert Septicémie (**SE_SES**).

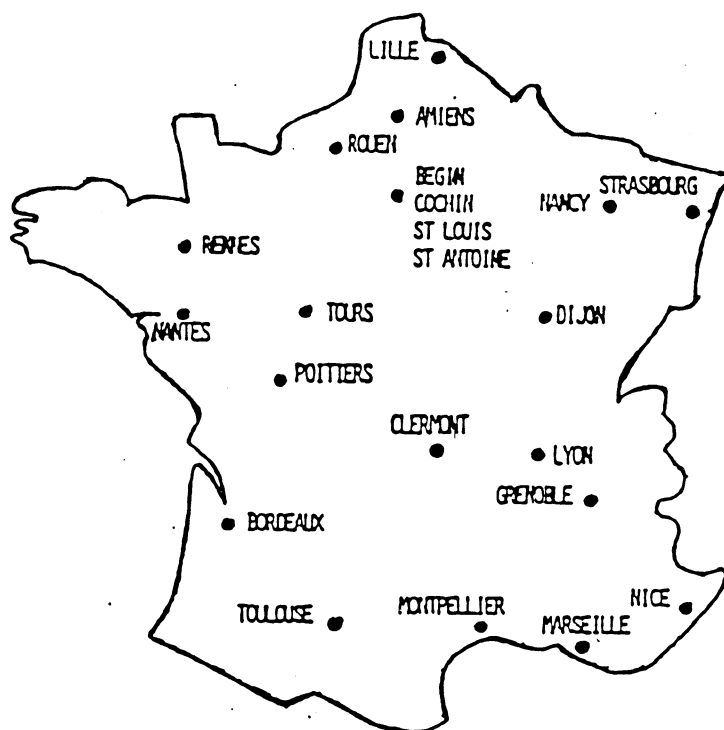


Figure 2.1 : Les Centres du Groupe SES

The S.E.S. network

(Ligue Française pour la prévention
des maladies infectieuses)

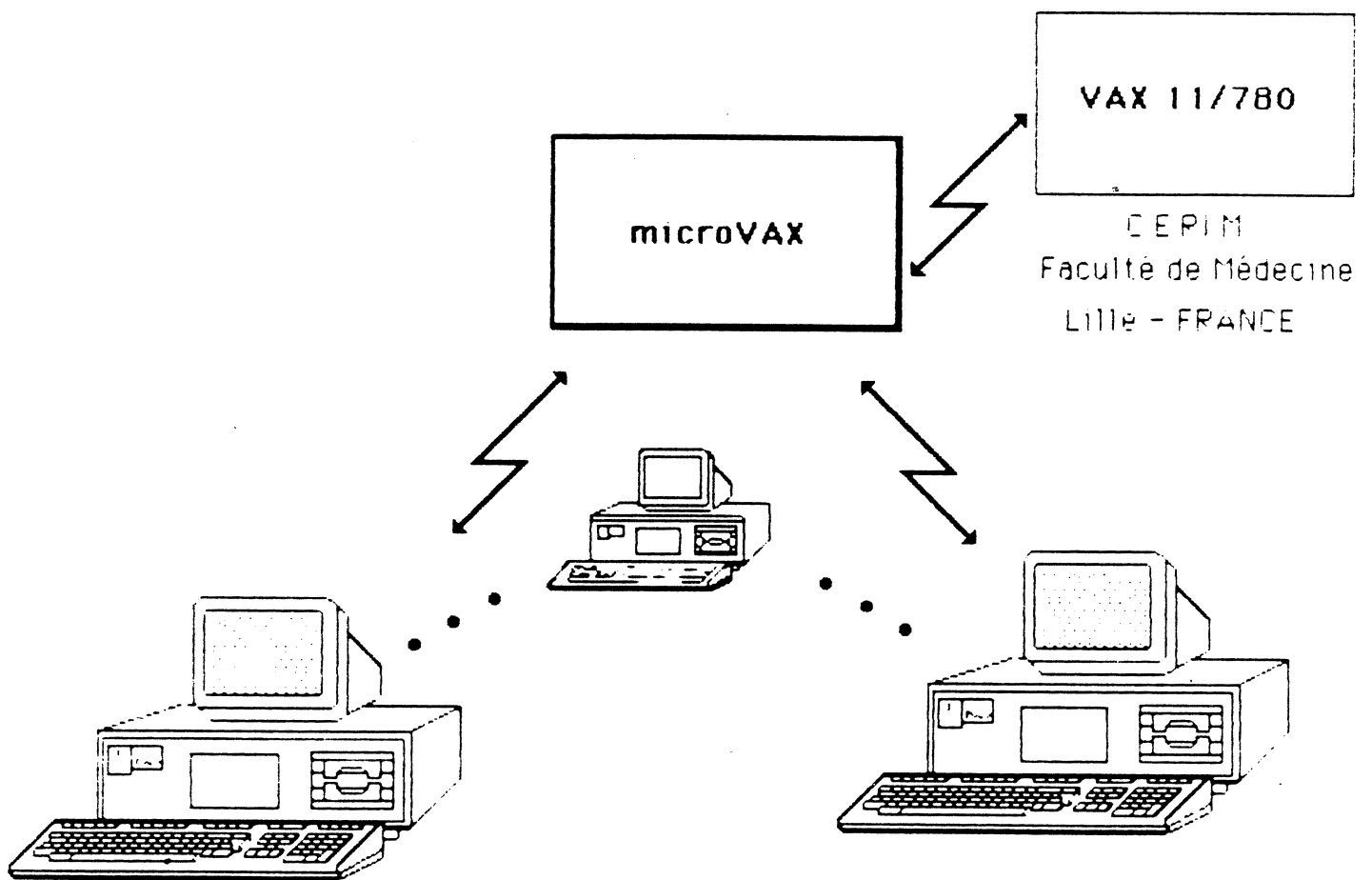


Figure 3.1

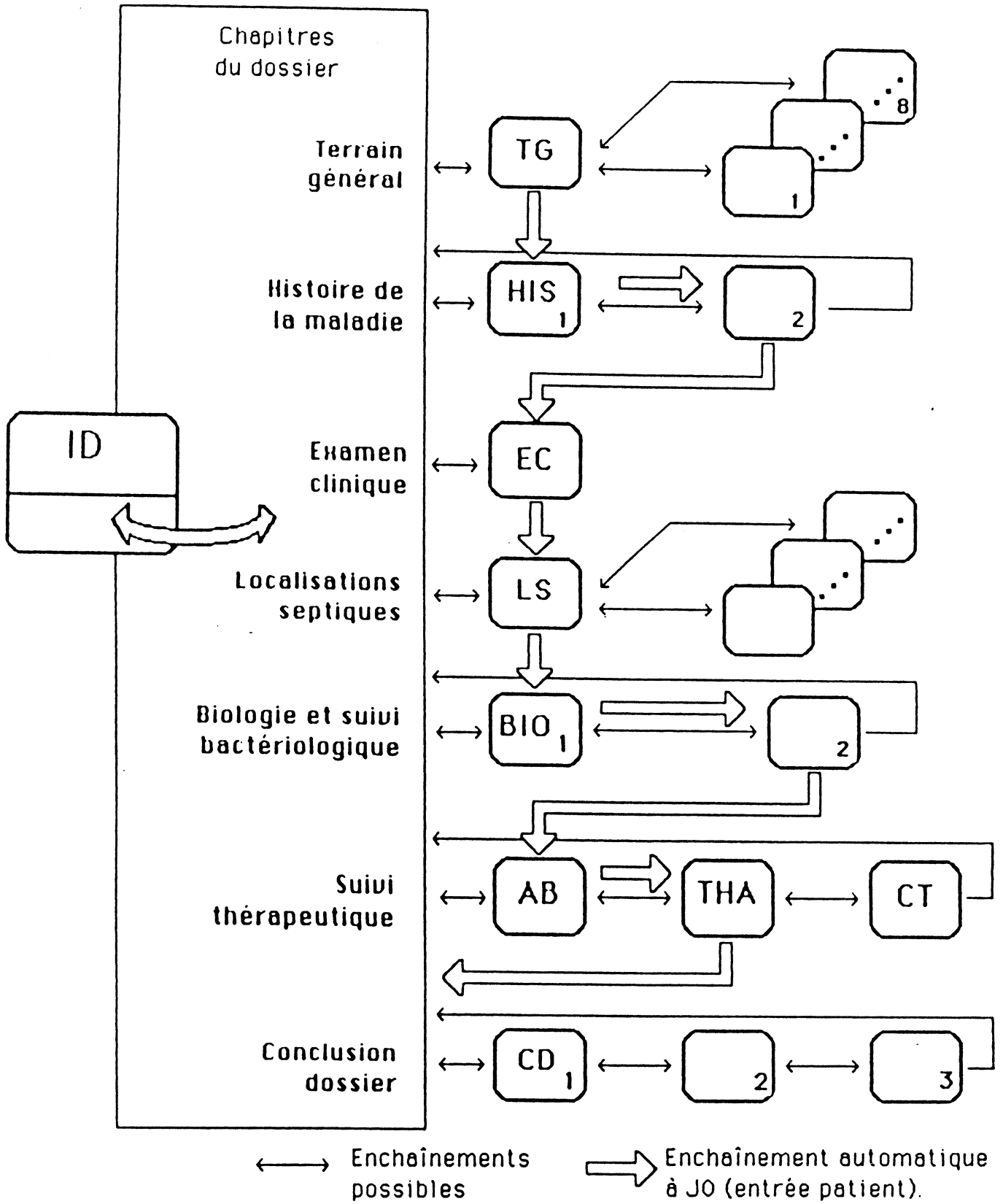
3 LA BASE DE DONNEES BD_SES

La base de données BD_SES a été réalisée dans le double objectif d'améliorer les connaissances de l'épidémiologie des septicémies et d'obtenir une importante base de cas pour valider le système expert SE_SES.

La méthodologie suivante a été employée :

- a. Définition en commun par les experts des 19 centres participant au projet SES de la structure du dossier informatisé.
- b. Développement du logiciel sur un matériel commun (Rainbow de Digital) à l'aide du logiciel de gestion de fichiers "Multilog".
- c. Saisie et exploitation locale des données à l'aide du micro-ordinateur et du logiciel implanté dans chaque centre.
- d. Constitution d'une base de cas nationale, obtenue en regroupant les observations de chaque centre.
- e. Retour de l'information vers chaque centre du groupe SES : Les résultats des principales études épidémiologiques réalisées sur les données centralisées ont été mis à la disposition des centres par le réseau télématique.

Cette base de données constitue donc un réseau épidémiologique (FIGURE 3.I). La démarche utilisée, association de chaque centre à la maîtrise d'oeuvre, possibilité d'exploitation locale et diffusion dans chaque centre des résultats des études globales, a permis une participation active de chacun de ces centres, et le recueil d'environ 450 septicémies par an depuis 1985.



Structure de la base de données S.E.S.

Figure 3.11

3.1 Structure du dossier

La FIGURE 3.II montre la structure générale du dossier BD_SES. Ce dossier est hiérarchisé en huit chapitres superposables à la démarche clinique habituelle : identification (IDSES), antécédents (TGSES), histoire de la maladie (HISSES), examen clinique (ECSES), localisations septiques (LSSES), résultats biologiques (BIOSES), thérapeutiques et antibiotiques (THQSES), conclusion de dossier (CDSSES).

La sélection d'un dossier se fait à partir du chapitre IDSES, qui contient l'état civil du patient. Ce chapitre est le seul à comporter des informations non anonymes. Ces dernières ne sont pas transmises lors de l'étape de centralisation. A partir de IDSES, un menu permet de sélectionner l'un quelconque des autres chapitres. Certains de ces chapitres contiennent eux-mêmes des sous-chapitres. Cette organisation confère à BD_SES une structure arborescente.

Pour les chapitres ECSES, LSSES, BIOSES et THQSES, une dimension chronologique a été prévue, les données qu'ils contiennent étant susceptibles d'évoluer pour un même patient (par exemple, changement d'antibiotiques au cours de l'hospitalisation).

La saisie des données a été facilitée par l'utilisation d'écrans de saisie, le contrôle actif de validité et de cohérence, l'affichage de messages explicatifs, l'enchaînement automatique de pages écran ou de chapitres. Pour les variables comportant un grand nombre de codes différents (antibiotiques et germes), une aide au codage 'en ligne' a été intégrée. Un certain nombre de fonctions utilitaires ont été programmées (autoformation, sauvegarde des dossiers sur disquettes..). Une brochure explicative très claire et concise a été éditée par le

CERIM et distribuée dans chaque centre.

3.1.1 Choix des items

Nous précisons ici les principaux facteurs qui ont guidé le choix de ces items par les experts du groupe SES.

- D'une manière générale, les items reconnus comme significatifs pour la prédiction du micro-organisme de la septicémie ont été inclus dans le dossier.
- Pour les données cliniques et les localisations septiques, les items figurant dans le dossier sont essentiellement d'ordre syndromique (par exemple "endocardite") et non sémiologique ("douleur thoracique"). Les seuls éléments sémiologiques retenus sont ceux du syndrome infectieux, tels que "fièvre", "sueurs", "purpura"...
- Le chapitre BIOSES était initialement très lourd (dans la version 1985 du dossier) : il regroupait la bactériologie, le bilan standard, l'immunologie et le bilan urinaire. L'exploitation statistique et épidémiologique des données de 1985 n'ayant pas mis en évidence de résultats significatifs pour de nombreux items, ce chapitre a fait l'objet en 1986 d'une importante simplification. Il regroupe désormais des données hématologiques (numération formule...), inflammatoires (vitesse de sédimentation...) et les résultats bactériologiques.

La base de données BD_SES comporte plus de 600 items. Naturellement, pour un patient donné, le volume des informations enregistrées est beaucoup plus restreint et la saisie d'un dossier peut généralement être effectuée en quelques minutes, compte tenu des nombreuses facilités offertes à l'utilisateur. Ce dossier informatisé réalise une description précise et complète des septicémies observées dans les différents centres, permettant ainsi une exploitation statistique et épidémiologique.

3.1.2 Critères de sélection des patients

Pour notre étude, sept critères d'inclusion des patients, répertoriés dans le chapitre CDESES, ont été retenus par les experts du groupe SES :

- Les septicémies à hémocultures positives.
- Deux localisations septiques non contigues et du même germe.
- Une infection d'un foyer ferme.
- Endocardite à hémocultures négatives.
- Brucellose, typhoïde, leptospirose, prouvées sérologiquement.
- Mycoplasme, ornithose, legionnella, prouvées sérologiquement.
- Choc toxi-infectieux à hémocultures négatives.

Ces critères définissent très précisément les septicémies du groupe SES. Environ 1500 observations ont ainsi été enregistrées pour les années 1985, 86, et 87.

3.2 Recrutement

3.2.1 La définition des septicémies du groupe SES.

Dans le domaine de la septicémie, les critères de sélection des patients varient selon les études. Cette variabilité est due essentiellement à deux facteurs :

- a. La fiabilité des résultats bactériologiques n'est pas absolue : Les hémocultures peuvent être positives chez des patients ne souffrant pas d'une septicémie. De tels résultats sont consécutifs à une contamination au moment du prélèvement sanguin ou en laboratoire [MACGREGOR 72]. Cependant, la mise en évidence d'au moins deux hémocultures positives du même germe associées à des signes cliniques de septicémie est généralement considérée comme suffisante pour lever le doute [MACGREGOR 72, WEINSTEIN 83a 83b]. Symétriquement, les résultats bactériologiques peuvent être négatifs chez des patients présentant les signes cliniques de la septicémie et devant être traités en conséquence : choc toxi-infectieux à hémocultures négatives par exemple.
- b. Il existe des divergences d'interprétation dans la définition de la septicémie : certains auteurs excluent de leur étude les patients présentant un état septicémique à point de départ lymphatique (typhoïde, brucellose), les salmonelloses ou les pneumopathies à hémocultures positives.

Nos critères de sélection définissent de manière très précise la nature des septicémies du dossier BD_SES. Ils prennent en compte les données cliniques et bactériologiques comme dans [MCGOWAN 75]. Ils peuvent être considérés comme optimaux en comparaison à d'autres études qui ne retiennent que les patients présentant un hyperthermie chez lesquels deux hémocultures au moins ont été positives pour le même germe [BABEAU 85], [SWANBON 79].

3.2.2 Les observations archivées dans BD_SES.

Les patients enregistrés dans la base de données proviennent des 19 centres de maladies infectieuses répartis sur l'ensemble du territoire français (FIGURE 2.I). Grâce à cette organisation en réseau, les observations recueillies constituent un reflet assez fidèle des septicémies observées en France aujourd'hui. Le nombre des patients recrutés à ce jour (plus de 1400), dépasse celui des séries récentes les plus importantes.

Certains points doivent cependant être précisés :

- a. Les centres du groupe SES sont des services spécialisés de maladies infectieuses de centres hospitalo-universitaires (CHU). Notre recrutement est de ce point de vue différent de celui des hôpitaux généraux.
- b. Les malades du groupe SES ont été suivis essentiellement en services de médecine (80%) et de réanimation (12%).
- c. Le TABLEAU 3.I résume les caractères généraux de notre population, obtenus à partir de 1330 dossiers exploitables. Il amène les remarques suivantes :

- Notre population est en majorité une population d'adultes.

- Le lieu d'acquisition : les fréquences des différents germes sont différentes selon le lieu d'acquisition de l'infection [MCGOWAN 79, WEINSTEIN 83a et 83b]. De plus, les septicémies nosocomiales (contractées à l'hôpital) sont généralement les plus mortelles [BEYTOUT 88, WEINSTEIN 83a et 83b]. La proportion de septicémies nosocomiales est différente selon les études : 50% dans la série de MCGOWAN, 66% dans celle de WEINSTEIN, 22% pour SES. Elle dépend de la nature des centres (CHU ou hôpitaux généraux), la population suivie dans chacun de ces centres étant différente [WEINSTEIN 83a et 83b]. Parmi les septicémies nosocomiales du groupe SES, 43% ont été contractées en service de médecine, 15% en réanimation et 42% en chirurgie. Cependant, notre population est peu concernée par la chirurgie lourde (transplantation) et la gynécologie (2,5%).

- Le terrain général joue un rôle important dans l'apparition d'une septicémie et dans la gravité du pronostic [BABEAU 85], [WEINSTEIN 83a 83b], [MAC CABE 83]. Les données apparaissant dans le TABLEAU 3.I montrent que notre population comporte un nombre élevé de patients présentant un tableau clinique sévère.

CARACTERES	FREQUENCES (%)	CARACTERES	FREQUENCES
AGE		TERRAIN GENERAL	
<15 ans	3%	Cancer	12%
16 à 30 ans	12%	Dysimmunité	10%
31 à 60 ans	31%	Cardio-vasculaire	33%
61 à 75 ans	30%	Pleuro-pulmonaire	10%
> 75 ans	24%	Digestif	17%
SEXE		Uro-génital	15%
Femme	41%	Neurologique	12%
Homme	59%	Diabète	12%
ACQUISITION		Rhumatologie	7%
Nosocomiale	22%	Toxique	14%
Communautaire	78%	Antécédent-chirurgie	8%
PORTES D'ENTREE		Dermatologie	6%
Cathéter	6%	Antécédent infectieux	12%
Urinaire	18%	LOCALISATIONS SEPTIQUES	
Digestive	10%	Cutanée	12%
Pulmonaire	9%	Stomatologique	4%
Cutanée	9%	Pleuro-pulmonaire	17%
ORL	4%	Cardio-vasculaire	11%
		Neuro-méningée	9%
		Ostéo-articulaire	8%
		Urogénitale	19%
		Abdominale	8%
		Hématologique	<1%
		Purpura	3%

TABLEAU 3.I : CARACTERES GENERAUX

La répartition des germes que nous obtenons sur 1330 septicémies (TABLEAU 3.II) est comparable à celles rapportées par d'autres études portant sur des séries importantes : un recensement effectué en 1984 par le laboratoire national de la santé auprès de nombreux laboratoires français [GOULET 86], destiné à préciser la fréquence et les circonstances d'isolement des bactéries dans les hémocultures, montre des prévalences analogues avec, en particulier, une prépondérance de l'Escherichia-coli (23,4%) et du Staphylocoque doré (15,4%) (respectivement 23% et 18% dans notre étude). Cette

prépondérance est également retrouvée par WEINSTEIN [WEINSTEIN 83a et 83b] sur une série de 500 septicémies recrutées entre Juillet 1975 et Mars 1977 au centre hospitalo-universitaire de Denver (Colorado).

GERMES	SES	GOULET	WEINSTEIN
Escherichia-coli	23%	23,4%	18,6%
Staphylocoque doré	18%	15,4%	13,4%
Streptocoque (A,D,NG,B)	12%		11%
Pneumocoque	11%		10%
Klebsielle	4%		6%
Pseudomonas	3%		5,5%

TABLEAU 3.II :FREQUENCES DES GERMES

3.3 Etudes épidémiologiques

Des études épidémiologiques ont été réalisées au CERIM sur les septicémies observées en 1985, 1986 et 1987. Nous ne détaillons pas ici ces études qui ont fait l'objet de nombreuses publications : outre une analyse descriptive globale [PORTIER 87], elles ont concerné en particulier les facteurs de mortalité [BEYTOUT 88] et [STHAL 87] , les septicémies nosocomiales [LACUT 86], les septicémies à portes d'entrée urinaire et pulmonaire [MASSARI 86] et [ROUE 86], les localisations septiques au cours des septicémies [BROTTIER-MANCINI 86]. Ces études sont actuellement poursuivies et le seront également dans les années qui viennent. Elles permettront d'obtenir des résultats de plus en plus significatifs au fur et à mesure de l'augmentation de la taille de notre échantillon.

Une synthèse des principaux résultats épidémiologiques extraits de ces différentes études a été mise à la disposition des centres du groupe SES par le réseau télématique. Les informations proposées sont regroupées dans la FIGURE 3.IV

MENU PRINCIPAL

Germes en fonction:

- de l'épidémiologie
- d'un terrain général
- d'une porte d'entrée
- d'une localisation

Epidémiologie en fonction:

- d'un germe
- d'un terrain général
- d'une porte d'entrée
- d'une localisation

Terrains généraux en fonction:

- d'un germe
- de l'épidémiologie
- d'une porte d'entrée
- d'une localisation

Portes d'entrée en fonction:

- d'un germe
- de l'épidémiologie
- d'un terrain général
- d'une localisation

Localisation en fonction:

- d'un germe
- de l'épidémiologie
- d'un terrain général

FIGURE 3.IV

Il est ainsi possible d'obtenir quasi instantanément, par exemple, le germe en fonction de la porte d'entrée (FIGURE 3.V). La deuxième colonne indique les fréquences des germes observés dans notre population, toutes septicémies confondues.

GERMES EN FONCTION DE LA PORTE
D'ENTREE: urinaire

		% GLOBAL
Escherichia Coli	70%	23%
Klebsielle	10%	4%
Proteus Indole -	7%	2%
Strepto. D Entérocoque	5%	3%
Proteus Indole +	4%	2%
PS. Aerugin. Carb. R	4%	2%
Staphylo Doré METI S	2%	15%
Enterobacter	2%	2%
Serratia	2%	1%
PS. Aerugin. Carb. S	2%	1%
Salmo. Non Thyphiques	1%	3%
Staphylo Doré METI R	1%	2%
Acinetobacter	1%	1%
Pneumocoque	0	11%
Salmo. Thyphiques	0	4%
Strepto. non groupable	0	4%
Staphylo. Blanc	0	3%
Meningocoque	0	3%
Autres germes	0	14%

FIGURE 3.V

4 LE SYSTEME EXPERT SEPTICEMIE (SE_SES)

Le rôle clinique du système expert et ses utilisateurs potentiels ont été définis dans notre introduction. Nous exposons dans ce paragraphe la méthodologie adoptée pour construire ce système expert.

4.1 La partition en deux modules

En cas de Septicémie, le traitement est fonction du germe responsable de l'infection et du patient. La démarche de l'expert en infectiologie et en antibiothérapie comporte deux étapes :

- a. Diagnostic du germe : l'expert se fonde sur un ensemble de signes cliniques et des données biologiques standard pour déterminer le ou les germe(s) le(s) plus probable(s).
- b. Traitement adapté au patient : la connaissance du germe pathogène ne résoud pas tous les problèmes. Le traitement dépend également de la nature du site infectieux et de l'état général du patient : tel antibiotique, "Céphalosporine de première génération", actif contre tel germe, "Pneumocoque", peut se révéler adapté à telle localisation septique, "Pleuro-pulmonaire" , mais sans effet pour une autre localisation septique, "Neuro-méningée" ; une septicémie due à un "Escherichia Coli" est généralement traitée par une association "Penicilline-Aminoside", mais lorsqu'il s'agit d'une femme enceinte, la deuxième molécule est contre-indiquée. De plus, les posologies, la durée du traitement et le mode prescription (mono ou bi-antibiothérapie) dépendent de la gravité de l'infection et de l'urgence thérapeutique.

Lorsque les résultats bactériologiques sont connus, souvent après 24 ou 48 heures, un antibiogramme est généralement fourni avec les résultats du laboratoire, qui définit les molécules actives contre le germe. Là encore, se pose le problème de la détermination du traitement adapté au patient. L'étape b) doit donc être réalisée.

Cette partition en deux problèmes complexes justifie le choix, opéré par les experts, de diviser le Système Expert Septicémie SE_SES en deux modules :

- **DIAGSES** pour l'aide au diagnostic du germe responsable de l'infection,
- **ABSES** pour l'aide à l'institution du traitement adapté au patient et au germe causal.

4.2 Diagnostic et traitement adapté : à partir de quelles données ?

Dans le cadre de notre application, quelles sont les données à partir desquelles le diagnostic et le traitement peuvent être obtenus ?

La nature de ces données dépend d'abord des utilisateurs potentiels du système expert. SE_SES a été conçu pour apporter une aide réelle au médecin non spécialiste en infectiologie et antibiothérapie, confronté à une septicémie. SE_SES n'est donc pas un outil sémiologique. Il s'adresse à des médecins compétents, capables de faire un diagnostic. Les informations sémiologiques, telles que "raideur à la nuque", n'ont pas été prises en compte dans l'élaboration de la base de connaissances, contrairement aux syndromes médicaux tels que "méningite".

De ce point de vue, les faits que l'on peut fournir au système ont été définis comme étant ceux qui permettent aux experts de déduire les probabilités bactériennes et d'instituer une antibiothérapie adaptée : ce sont les seules données de l'examen clinique et de la biologie standard. Ces éléments appartiennent à quatre familles principales :

- a. Le terrain général du patient : cancer, immuno-dépression...
- b. Les localisations septiques primaires et secondaires : pulmonaires, digestives...
- c. L'allure clinique de l'infection : en particulier le lieu d'acquisition de l'infection (communautaire ou nosocomiale) et les seuls éléments sémiologiques qui sont ceux du syndrome infectieux (fièvre, choc infectieux)...
- d. Les résultats du premier bilan biologique : globules blancs, vitesse de sédimentation...

Toutes ces informations peuvent être obtenues lors de l'examen clinique réalisé par le médecin ou quelques heures après pour les paramètres biologiques. Ces faits sont identiques ou peuvent être déduits à partir des items du dossier informatisé BD_SES.

4.3 Le choix du formalisme de représentation des connaissances.

BUCHANAN [BUCHANAN 84] indique trois points essentiels qui doivent être vérifiés pour un bon modèle de représentation des connaissances :

- a. **Extensibilité** : le modèle de représentation des connaissances doit être suffisamment souple pour permettre une modification aisée de la base de connaissances (validation, maintenance).
- b. **Simplicité** : les concepts utilisés ne doivent pas être trop complexes de façon à permettre à l'expert non informaticien de transcrire aisément son savoir.

- c. **La connaissance doit être explicite** pour faciliter la recherche des erreurs (validation) et la justification des réponses du système (validation et utilisation).

Les chercheurs en Intelligence Artificielle se sont attachés à respecter ces principes dans la mise au point de différents modèles de représentation des connaissances [RICH 85] : règles de production avec objets simples, réseaux sémantiques, schémas ou objets structurés, langages orientés objets. On trouvera dans VOYER [VOYER 85] ou FIESCHI [FIESCHI 88] un exposé introductif de ces différents formalismes.

Comme le souligne REGGIA [REGGIA 85], il n'existe pas, dans l'absolu, de meilleur modèle de représentation des connaissances, mais le choix du formalisme devrait prendre en compte trois facteurs :

1. le pouvoir d'expression du formalisme : celui-ci doit être aussi proche que possible de la représentation naturelle de l'expert.
2. La nature du problème à résoudre : pour tel problème, à caractère hautement déductif, on pourra utiliser de préférence les règles de production alors que, pour tel autre problème, où l'on doit modéliser un grand nombre de connaissances statiques, il sera plus adéquat d'utiliser des objets structurés.
3. Le degré de dépendance du problème vis-à-vis du contexte : il est difficile d'utiliser des règles de production si les prémisses de chaque règle doivent décrire une grande quantité de données.

Nous avons choisi en 1985, date du début de notre projet, le moteur d'inférence "SAM" [GASCUEL 81]. SAM est un moteur "d'ordre 0" (sans variable) qui permet de gérer des règles de production manipulant des objets simples. Les règles peuvent être pondérées par des coefficients de vraisemblance, pour prendre en compte l'incertitude du raisonnement. C'est un système monotone (il n'y a pas de retrait de connaissance au cours du raisonnement) qui fonctionne essentiellement en "chaînage avant".

Ce choix se justifie principalement par les arguments suivants :

- a. Le diagnostic et la thérapeutique de l'expert en infectiologie et en antibiothérapie dépendent d'un faisceau d'arguments utilisant chacun un nombre limité de signes cliniques et biologiques. La démarche intellectuelle qu'il utilise est à caractère hautement déductif.
- b. Dans le cadre de notre problème, le diagnostic et la thérapeutique, tout fait affirmé n'est jamais remis en cause, un malade présentant les mêmes symptômes avant et après la consultation. Le raisonnement est donc monotone.
- c. Face à une septicémie, la prise de décision diagnostique et thérapeutique est réalisée par un raisonnement sous incertitude. Le formalisme des coefficients de vraisemblance permet de prendre en compte cette incertitude [SHORTLIFFE 75].
- d. Le but à atteindre est la détermination du germe causal et de la bonne antibiothérapie, inconnus a priori. Le "chaînage avant" répond bien à cette situation.

- e. La syntaxe particulièrement simple avec laquelle les connaissances ont pu être formalisées, a permis aux experts eux-mêmes de transcrire leurs connaissances. Il s'agit, comme le souligne WEISS [WEISS 85], d'un avantage non négligeable. Beaucoup de problèmes relevant de l'expertise sont en effet trop complexes pour que les connaissances mises en jeu puissent être formalisées par l'intermédiaire d'un non expert du domaine. Pour notre application, le recueil d'une expertise multiple a ainsi été facilité.
- f. SAM a par ailleurs été mis au point essentiellement pour les applications médicales. Les propriétés intéressantes qui en découlent seront exposées au paragraphe 4.7.

4.4 la méthodologie d'acquisition des connaissances

La base de connaissances de SE_SES (DIAGSES et ABSSES) a été élaborée grâce à la collaboration étroite de 40 experts issus de 19 centres français de maladies infectieuses constituant le réseau SES. Cette organisation en réseau a permis de construire une base de connaissances à partir de compétences multiples et de points de vue différents sur les problèmes à résoudre. L'intérêt d'une telle démarche a été souligné en particulier par REGGIA et WEISS ([REGGIA 85, [WEISS 85]).

La méthodologie que nous avons employée a favorisé la collaboration de ces experts pour construire une base de connaissances complète, précise et cohérente, même lorsqu'il existait des divergences sur certains points diagnostiques ou thérapeutiques [BEUSCART 88a].

Etape I :

Le groupe SES a été divisé en quatre secteurs correspondant à des impératifs géographiques : le Nord Ouest (Lille), l'Est (Nancy), le Sud-Ouest (Toulouse) et le Sud-Est (Grenoble). Dans chaque région, un médecin a été désigné comme coordonnateur du projet dans son secteur géographique. Chaque groupe devait rédiger des règles correspondant à un domaine particulier de la base de connaissances à élaborer.

Cette connaissance peut s'exprimer selon deux points de vue différents :

1. Les septicémies causées par un germe donné : septicémie à *Escherichia Coli* par exemple.
2. Les signes cliniques liés à la septicémie : septicémie et immuno-dépression, septicémie et localisations pulmonaires...

Nous avons conservé cette expression duale de la connaissance pour construire les règles d'inférence. Par exemple, un groupe a été chargé de rédiger les règles concernant les septicémies à *Escherichia Coli*, un autre groupe celles concernant les septicémies à localisations septiques urinaires. Naturellement, des règles étaient communes et la collaboration des deux groupes a permis de confronter et de prendre en compte les deux points de vue.

Etape II :

Les experts de chaque secteur géographique se sont réunis mensuellement pour opérer un premier tri critique et aboutir à un consensus. Ensuite, les connaissances transcrites ont été soumises aux autres groupes pour une évaluation plus globale.

L'ensemble du groupe SES s'est ensuite réuni trimestriellement. Les règles ont alors été revues une par une, thème par thème. C'est de cette confrontation générale, après agrément de l'ensemble des experts du groupe SES, qu'est née la première version de la base de connaissances.

Etape III :

20 médecins du groupe SES se sont alors réunis pour tester les performances du système expert sur un ensemble de cas cliniques bien documentés. A l'examen des réponses du système, des corrections ont alors été réalisées : suppression de règles et modifications de règles, ajustements de coefficients de vraisemblance. C'est à l'issue de cette étape que les premières évaluations purent commencer (paragraphe 4.10).

4.5 L'expertise

L'expertise distingue deux types de connaissances :

Les connaissances descriptives : ce sont des connaissances statiques, ne faisant pas progresser le raisonnement. Elles ont pour but de définir et structurer les connaissances du domaine. Par exemple, "myélome et maladie de Waldenström sont des dysglobulinémies malignes".

Les connaissances dynamiques : ce sont les règles d'inférence proprement dites. La plupart de ces règles sont pondérées par des coefficients de vraisemblance (CV). Elles expriment des connaissances expertes de type heuristique. Ces règles de production qui traduisent le raisonnement incertain sont du type :

"Si PREMISSE ALORS CONCLUSION <CV>".

La prémisse peut être un fait ou une association de faits. Par exemple,

"Devant ce patient présentant une méningite associée à un purpura, il est très vraisemblable que le germe responsable de la septicémie soit un méningocoque" s'exprime par la règle :

"SI MENINGITE ET PURPURA ALORS MENINGOCOQUE <0.80>".

La conclusion peut être négative. La règle,

"SI DERIVATION DU LCR ET MENINGITE ALORS NON PNEUMOCOQUE <0.50>",

traduit la connaissance experte : "si le patient présente une méningite et une dérivation du LCR, alors le germe en cause n'est probablement pas (0,5) un Pneumocoque".

Les coefficients de vraisemblance manipulés par SAM sont proches de ceux de MYCIN [SHORTLIFFE 75], comme nous le verrons au paragraphe 4.7. Le CV traduit le degré de confiance que l'expert accorde à l'affirmation contenue dans la règle.

L'usage de ces coefficients de vraisemblance soulève des problèmes théoriques, comme nous l'avons vu au chapitre I. Ils ont cependant montré leur efficacité dans des systèmes reconnus, tels que MYCIN, ONCOCIN et PUFF. Du point de vue pratique, il est clair que l'évaluation correcte de ces coefficients est une condition nécessaire à l'obtention d'une base de connaissances pertinente. Nous reviendrons sur ce point dans la suite de notre exposé.

Lors de l'étape d'acquisition des connaissances, une démarche pragmatique a été adoptée. Pour toute règle incertaine du type "Si Condition alors Conclusion <CV>", les experts ont évalué le CV de la manière suivante :

CV = +0,8 si la conclusion est très vraisemblable,
= +0,5 si elle est vraisemblable,
= +0,2 si elle est peu vraisemblable, mais possible.

SAM permet également de réaliser un raisonnement en logique pure (paragraphe 4.7). En effet, certaines règles expertes expriment une connaissance certaine :

"SI GANGRENE GAZEUSE ALORS ANAEROBIE".

Le système expert, comme nous l'avons souligné, suppose que l'examen clinique a été réalisé par un médecin. Il raisonne donc à un niveau élevé de connaissances. Par exemple, il n'existe pas, dans la base de connaissances, de règles qui permettent de déduire "méningite" à partir de l'association "céphalées, diarrhée, photophobie, nausées et vomissements". Ce diagnostic de "méningite" est du ressort du médecin.

Les règles pondérées par des coefficients de vraisemblance permettent d'évoquer, face à un contexte clinique, les conclusions les plus probables, mais aussi de raisonner en terme d'utilité diagnostique ou thérapeutique. Considérons un exemple : une septicémie déclarée au cours d'une grossesse est souvent causée par l'Escherichia Coli. Cependant, il est possible, dans ce contexte, de retrouver une Listeria, bactérie ayant une résistance beaucoup plus importante aux antibiotiques. Il est donc nécessaire, pour la conduite à tenir et le traitement d'une septicémie au cours de la grossesse, de penser à cette possibilité. Les experts ont introduit dans la base de connaissances des règles d'utilité comme celle correspondant à notre exemple :

"SI GROSSESSE ALORS LISTERIA <0.30>".

DIAGSES regroupe les règles destinées principalement à l'identification du (des) germe(s) responsable(s) de l'infection. Mais **DIAGSES** permet également de déterminer les examens complémentaires utiles ou nécessaires.

Les règles de **DIAGSES** sont hiérarchisées de façon superposable au raisonnement des experts : partant des signes cliniques (faits initiaux) pour déduire des diagnostics intermédiaires, de ces diagnostics intermédiaires pour évoquer les germes, des germes pour proposer des examens complémentaires.

Par exemple :

"SI GROSSESSE ALORS INFECTION URINAIRE <.5>",
puis "SI INFECTION URINAIRE ALORS ESCHERICHIA COLI <.8>",
et "SI ESCHERICHIA COLI ET RETENTION_BILIAIRE
ALORS ECHOGRAPHIE_ABDOMINALE".

Les connaissances de **ABSES** permettent de fournir les conseils thérapeutiques. Les faits initiaux sont les éléments du tableau clinique du patient nécessaires à la prescription d'une antibiothérapie adaptée : localisation septique, terrain général et germe(s) responsable(s) de l'infection.

Les conseils thérapeutiques fournis par **ABSES** comportent plusieurs points :

- la liste des antibiotiques conseillés par les experts,
- le mode de prescription : mono ou bi-antibiothérapie,

- les contre-indications éventuelles, la posologie et les durées de traitements.

4.6 Principe d'utilisation du système

La partition en deux modules, DIAGSES et ABSES, permet d'utiliser le Système Expert Septicémie de deux façons :

- a. En situation d'urgence, lorsque les résultats bactériologiques ne sont pas encore connus. Deux étapes sont alors exécutées séquentiellement :

1) **DIAGSES** : à partir des signes cliniques et des paramètres biologiques que le médecin fournit au système, celui-ci déduit les germes les plus probables et propose si nécessaire des examens complémentaires. DIAGSES propose un, deux, le plus souvent trois germes possibles, par ordre décroissant de vraisemblance.

2) **ABSES** : immédiatement après l'exécution de DIAGSES, le médecin fournit à ABSES le germe reconnu comme causal. Les autres informations nécessaires à l'antibiothérapie adaptée sont conservées en mémoire.

- b. Lorsque les résultats bactériologiques sont connus, pour adapter l'antibiothérapie au patient. L'étape ABSES est alors réalisée directement. Il faut fournir au système le germe causal ainsi que les éléments cliniques nécessaires à l'institution d'une antibiothérapie adaptée : localisation septique, terrain général.

La FIGURE 4.I illustre le schéma d'interrogation du système expert SE_SES.

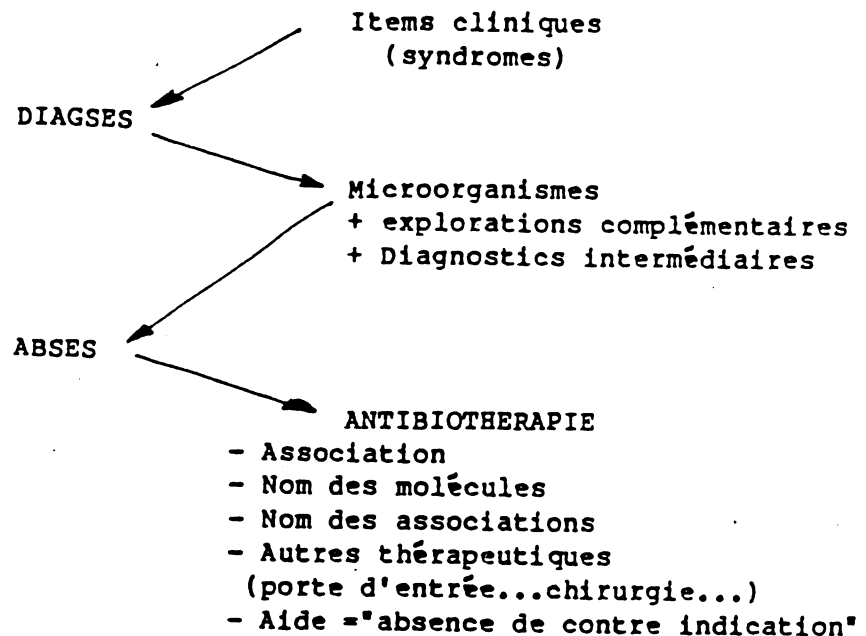


FIGURE 4.I

4.7 Caractéristiques du moteur d'inférence SAM

SAM est un moteur d'inférence mis au point par O. GASCUEL en 1981. Ce système est actuellement utilisé par d'autres équipes dans le cadre de systèmes experts destinés au traitement des hypertensions et des accidents vasculaires cérébraux. Nous en indiquons ici les caractéristiques principales. Pour une description plus précise, on consultera [GASCUEL 81].

4.7.1 Le langage de représentation des connaissances

Le langage de représentation des connaissances comporte des tautologies et des règles de production. Les objets manipulés par ce

langage sont répartis en quatre catégories sémantiques : signes et items valables, syndromes, diagnostics et thérapie, qui correspondent à des étapes du raisonnement médical.

On distingue deux types de tautologies :

1) **CONSTITUENT** : "F1 et F2 CONSTITUENT P". Cette tautologie peut être transcrite en logique des propositions par :

$(F1 \text{ OU } F2) \Leftrightarrow P.$

2) **SEXCLUENT** : "F1 et F2 SEXCLUENT", correspond en logique des propositions à :

$(F1 \text{ ET NON } F2) \text{ OU } (\text{NON } F1 \text{ ET } F2).$

Ces tautologies permettent une structuration a-minima de la connaissance.

On distingue deux types de règles de production :

1) $E(P_i)$ **IMPLIQUE** C. $E(P_i)$ est une expression, P_i et C des objets de la base de connaissances. La transcription en logique des propositions est : " $E(P_i) \Rightarrow C$ ". Ce type de règle permet d'exprimer des connaissances certaines, indiscutées.

2) $E(P_i)$ **EST EN FAVEUR** C **<CV>**. Ces règles permettent d'exprimer les connaissances expérimentales, incertaines.

Une distinction est donc faite entre deux types de connaissances : l'une indiscutée et l'autre incertaine. Ainsi, à un même objet, sont attachés trois coefficients. Le premier est un coefficient logique CL qui permet de suivre un raisonnement proche du calcul des propositions [DELAHAYE 87]. Les deux autres coefficients permettent de suivre un raisonnement incertain : l'un "en faveur" de l'objet (CVO pour CVOui), l'autre en "défaveur" (CVN pour CVNon).

C'est la combinaison de ces coefficients à travers le réseau des règles d'inférence, en liaison avec les coefficients de vraisemblance attachés aux règles (CV) qui permet de propager le raisonnement. Examinons comment SAM gère ces coefficients.

4.7.2 Les calculs sur les coefficients

On désigne classiquement par base de faits, l'ensemble des faits initiaux et des faits déduits par le moteur [VOYER 87].

a. Le coefficient logique CL d'un fait A est défini par :

$CL(A) = 1$ si le fait A est vrai, c'est-à-dire présent dans la base de fait,

$= 0$ si le fait A est faux, c'est à dire "NON A" présent dans la base de fait,

$= 1/2$ sinon.

b. Le coefficient résultant d'un fait CR est défini par :

$CR = CL$ si $CL = 1/2$,

$= (CVO-CVN+1)/2$ sinon.

c. L'évaluation du CR d'une expression est réalisée à l'aide des formules :

$CR(\text{NON } A) = 1-CL(A)$,

$CR(A \text{ ET } B) = \text{MIN}[CR(A), CR(B)]$.

d. Considérons la règle "A EST EN FAVEUR D <CV>", et posons

- $ARG(A) = CVO(A)-CVN(A)$ si $CL(A) = 1/2$,

$= 1$ (respectivement 0) si $CL(A) = 1$ (respectivement 0).

ARG(A) est "l'argument" (en faveur ou en défaveur) de A.

- $CVO_i(D)$ (respectivement $CVOf(D)$), valeur du coefficient "en

favor" de D avant (respectivement après) le déclenchement de la règle.

- L'opérateur SOM, défini par : $x \text{ SOM } y = x+y-xy$.

On a alors :

- La règle n'est pas déclenchée si $\text{ARG}(A) \leq 0,2$

- $\text{CVO}_f(D) = [\text{CVO}_i(D)] \text{ SOM } [\text{ARG}(A) \times \text{CVR}]$ si $\text{ARG}(A) > 0,2$.

e. Pour la règle "A EST EN FAVEUR NON D", il suffit de remplacer "CVO" par "CVN" dans les formules précédentes.

f. Considérons la règle "A IMPLIQUE D " :

- si $\text{CL}(A) = 1$ alors $\text{CL}(D) = 1$. On réalise donc un calcul des propositions sur les règles "IMPLIQUE" dont les prémisses sont certaines.

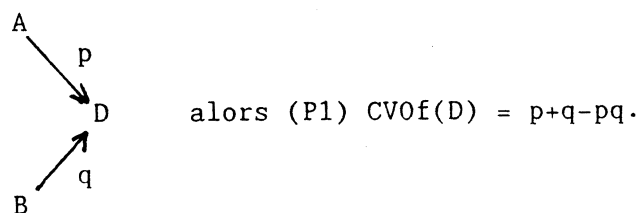
- Si le fait A est incertain, la règle est interprétée par le moteur comme une règle incertaine avec un CV égal à 1 ("A EST EN FAVEUR D <1>"). Le CR de D est alors calculé grâce aux formules précédentes.

g. Pour la règle "A IMPLIQUE NON D ", les calculs sont semblables, en remplaçant " $\text{CL}(D) = 1$ " par " $\text{CL}(D) = 0$ ", et "D" par "NON D".

On peut remarquer que, pour le raisonnement incertain, le mode de combinaison des coefficients de vraisemblance peut se résumer par deux combinaisons élémentaires :

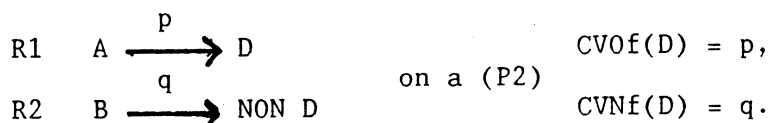
- Combinaison parallèle.

Considérons deux règles "EN FAVEUR" de D. A et B sont des faits vrais, D un fait inconnu avant l'activation de la règle ($\text{CVO}_i(D) = \text{CVN}_i(D) = 0$), p et q les poids des règles :



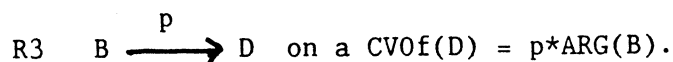
La formule est équivalente dans le cas deux règles "EN DEFAVEUR".

Pour le cas où deux règles sont contradictoires :



- Combinaison séquentielle.

B est un fait intermédiaire, tel que $ARG(B) = CVOf(B) - CVNf(B) = b$ ($b > 0.2$) avant le déclenchement de la règle et D est inconnu :



Les opérations élémentaires précédentes sont comparables à celles effectuées dans MYCIN. Dans EMYCIN [VAN MELLE 79], la formule de combinaison parallèle dans le cas de deux règles contradictoires (P2) a été modifiée de la manière suivante :

$$\text{Si } p > q, CVOf(D) = \frac{p-q}{1-\text{MIN}(|p|, |q|)}$$

La formule est analogue avec $CVNf(D)$ (et $q-p$) si $q > p$.

L'exemple suivant illustre le mode de propagation des coefficients de vraisemblance à travers le réseau des règles d'inférence :

Base de faits : A,B,C,D.

Base de règles :

- (R1) A \longrightarrow X <a>,
 (R2) B \longrightarrow X ,
 (R3) C \longrightarrow NON X <c>,
 (R4) D \longrightarrow Y <d>,
 (R5) X \longrightarrow Y <x>.

Inférence :

(R4) donne $CV0(Y) = d$,

(R1) et (R2) donnent $CV0(X) = a+b-ab$,

(R3) donne $CVN(X) = c$. Donc $ARG(X) = a+b-ab-c$.

Si $ARG(B) >.2$ alors (R5) est déclenchée :
 $CV0(Y) = d+x(a+b-ab-c)-dx(a+b-ab-c)$.

SINON (R5) n'est pas déclenchée : $CV0(Y) = d$.

SAM fonctionne en "chaînage avant" pour déduire les conclusions possibles. Une option "chaînage arrière" existe cependant, qui permet de sélectionner et de demander à l'utilisateur des informations supplémentaires dans le but d'améliorer la discrimination entre les conclusions fournies par le "chaînage avant".

L'avantage principal de SAM vis-à-vis de MYCIN est de mener en parallèle deux raisonnements, l'un logique et l'autre incertain, et de donner la priorité au raisonnement logique lorsqu'il est possible.

D'autre part, dans SAM, les règles sont préordonnées, de telle sorte que toute règle ayant un item "X" en conclusion, précède celles qui utilisent cet item "X" en prémisse. Cet ordre sur les règles permet au moteur de dérouler très rapidement le raisonnement.

On peut cependant formuler quelques critiques. Dans le cadre de notre application, la principale concerne la capacité du système à expliquer son raisonnement. On peut regretter que, contrairement à MYCIN qui dispose du module explicatif TEIRESIAS, SAM se limite à produire la trace des règles inférées. Pour d'autres applications, l'absence de méta-règles et d'un formalisme plus riche pour représenter des connaissances statiques complexes peut être pénalisante.

4.8 Interfaces utilisateurs

Pour exécuter SAM, il est nécessaire de connaître la syntaxe exacte des faits manipulés par la base de connaissances. D'autre part, les résultats qu'il fournit se présentent sous la forme d'un tableau peu lisible pour l'utilisateur non averti, répertoriant les faits initiaux et les faits déduits (en particulier les diagnostics) affectés de leurs coefficients de certitude. Cette situation n'était pas compatible avec une utilisation réelle du système expert (chapitre I).

Nous avons donc développé deux interfaces :

- a. Le premier module, **SYNONE**, est destiné à aider le médecin à introduire les faits initiaux utiles à SES. SYNONE se réfère à un dictionnaire de termes médicaux pour comprendre et corriger, lorsque c'est nécessaire, les informations saisies. Il vérifie par ailleurs que les informations indispensables à une bonne expertise lui ont bien été fournies. Le médecin a également la possibilité d'utiliser un système de menus pour transmettre les informations à SES.

- b. Le deuxième module, GLOSE [ROUSSEL 88], interprète en français correct les résultats du tableau complexe fournit par le Système Expert Septicémie . GLOSE est un générateur de langage naturel. Il est capable d'ordonner correctement les phrases, de réaliser les accords... DIAGSES propose généralement trois germes possibles, classés par ordre décroissant de vraisemblance. GLOSE affecte des adverbes selon les valeurs de ces coefficients (probablement, éventuellement...).

4.9 Les différences avec MYCIN

MYCIN [SHORTLIFFE 84] a également été mis au point pour le diagnostic et le traitement des maladies infectieuses et en particulier des septicémies.

Au contraire de SES, MYCIN utilise deux types d'expertise : l'une bactériologique et l'autre clinique. MYCIN ne prend en compte, au cours de l'étape diagnostique, que des informations bactériologiques. Il suppose que l'on dispose de renseignements précis sur le germe en cause, correspondant à des connaissances bactériologiques et non cliniques : site des hémocultures, forme du micro-organisme ("bâton",...), coloration, caractéristiques micro-biologiques. Ce n'est que lors de l'étape traitement, que les informations cliniques sont prises en compte. Cette démarche ne correspond pas à la démarche habituelle de l'expert infectiologue confronté à une septicémie. De notre point de vue, MYCIN semble ainsi mélanger deux type d'expertise.

En revanche, MYCIN dispose, avec le module TEIRESIAS [DAVIS 79], d'une capacité exceptionnelle à expliquer le raisonnement et d'une aide à la modification de la base de connaissances.

4.10 Evaluation initiale de SES

Notre étude a concerné l'évaluation de la justesse des réponses de SES en tant qu'outil clinique. Nous avons vu que le traitement d'une septicémie était conditionné par la connaissance du germe responsable de l'infection. Nous avons donc évalué :

- les performances diagnostiques de SES (DIAGSES),
- la pertinence de ses conseils thérapeutiques lorsque le germe réel est connu (ABSES). Dans le cadre de ce mémoire, nous n'exposerons pas les résultats obtenus sur ABSES que l'on pourra trouver dans [BEUSCART 88a].

Dans le cadre de notre application, le "Gold standard" est connu dès lors que l'on dispose des résultats des examens bactériologiques.

a. **Evaluation de la justesse "absolue" des réponses de DIAGSES.**

Les performances de DIAGSES ont été évaluées sur 115 cas cliniques réels, provenant des différents centres du groupe SES. Ces cas étaient tous très bien documentés : antécédents, signes cliniques, paramètres biologiques, résultats bactériologiques, thérapeutique. Généralement, seuls six ou sept signes ont été jugés utiles pour le diagnostic et fournis au système.

Nous savons que DIAGSES propose généralement trois germes possibles, par ordre décroissant de vraisemblance. Les réponses du système pour ces 115 dossiers ont été comparées aux germes

réels. Nous avons adopté l'échelle de valeurs suivante :

1) Lorsque le germe réel était proposé en première position par DIAGSES, il s'agissait d'une bonne réponse et nous avons attribué la note 1.

2) Lorsque le germe réel était cité dans l'une des trois premières positions, nous avons considéré qu'il s'agissait d'un bon avis et attribué la note 2. Ce type de réponse reproduit en effet l'avis d'un infectiologue : "devant cette septicémie à infection urinaire, je pense d'abord à l'Escherichia Coli (première position), sinon à la Klebsielle ou au Proteus (deuxième ou troisième position).

3) Dans les autres cas, nous avons considéré que la réponse était inexacte.

Les résultats obtenus apparaissent dans le TABLEAU 4.I

NOTE 1 (BONNE REPONSE):	59%
NOTE 2 (BON AVIS) :	86%
REPONSE INEXACTE :	14%

TABLEAU 4.I : EVALUATION SUR 115 DOSSIERS

Le détail des réponses par micro-organisme montre que l'Escherichia Coli, le Staphylocoque doré, le Pneumocoque et le Méningocoque sont bien identifiés par DIAGSES (TABLEAU 4.II). Ce sont les germes les plus fréquemment rencontrés dans les septicémies. Les autres germes sont identifiés en deuxième ou troisième position. Nous retrouverons une situation analogue au chapitre III B, Paragraphe 2 lorsque nous analyserons les résultats obtenus sur l'ensemble des dossiers de la base de données.

GERMES	NOTE I	NOTE II
Staph Doré	64%	86%
Staph Blanc	0	36%
Escherichia Coli	87%	98%
Pneumocoque	71%	83,5%
Méningocoque	10%	80%
Salmonelle	0	67%
Strepto A	37%	75%
Strepto D	10,5%	63,5%
Strepto NG	53%	60%
Klebsielle	0	10%
Pseudomonas	0	11%
Proteus	0	60%
Anaérobies	0	0
Candida	0	0

TABLEAU 4.II

b. Estimation de l'utilité de DIAGSES.

Afin de d'estimer l'utilité réelle du système expert, nous avons mis en compétition DIAGSES et des prescripteurs habituels d'antibiotiques, praticiens non experts en infectiologie et en antibiothérapie. Une étude de concordance entre les avis de ces différents acteurs (DIAGSES-Cliniciens) a été réalisée. On trouvera dans [CBEUSCART 88] les résultats complets que nous résumons ici.

40 cas cliniques, 20 considérés comme faciles, 20 comme difficiles, ont été fournis par des experts ne participant pas au projet SES. Ces cas cliniques ont été soumis d'une part à des médecins hospitaliers (chefs de clinique, assistants des hôpitaux), d'autre part au système expert. Les résultats obtenus sur les cas faciles, puis sur les cas difficiles, apparaissent dans le TABLEAU 4.III.

Les performances de DIAGSES sont comparables à celles obtenues sur les 115 dossiers précédents. Dans les cas faciles, on observe que les performances des praticiens et du système expert sont proches. Dans les cas difficiles, le système expert se montre en revanche nettement plus pertinent dans son diagnostic que les praticiens. Ces résultats sont en faveur de l'utilisation de SES.

	CAS FACILES	
	SES	NON-experts (moyenne)
NOTE 1 (BONNE REPONSE):	70%	58%
NOTE 2 (BON AVIS) :	100%	89%
REPONSE INEXACTE :	0	11%
	CAS DIFFICILES	
	SES	NON-experts (moyenne)
NOTE 1 (BONNE REPONSE):	50%	10%
NOTE 2 (BON AVIS) :	85%	18%
REPONSE INEXACTE :	15%	82%

TABLEAU 4.III

CHAPITRE III B: EVALUATION DU SE SES SUR LA BASE DE DONNEES

CHAPITRE III B : EVALUATION DU SE_SES SUR LA BASE DE DONNEES

1 TRANSFERT BASE DE DONNEES - SYSTEME EXPERT SE_SES.

Nous avons réalisé un logiciel, LINXES, qui permet de faire fonctionner le Système Expert Septicémie (SE_SES) sur l'ensemble des dossiers de la base de données BD_SES et d'évaluer les performances du système sur les cas de septicémie de base de données.

1.1 Faisabilité et objectifs

La base de données BD_SES permet le recueil de 450 dossiers par an depuis 1985, provenant des 19 centres participant au projet. Plus de 600 items cliniques et biologiques peuvent être enregistrés dans ce dossier informatisé, réalisant une description précise et complète des cas de septicémie observés. Le(s) germe(s) responsable(s) de l'infection (données bactériologiques) et le traitement suivi par le malade sont également enregistrés.

Les items de ce dossier informatisé ont été définis en commun par les experts des 19 centres du groupe qui ont réalisé par ailleurs les bases de connaissances de SE_SES. Il était d'emblée dans leurs intentions d'utiliser la base de données comme base de validation des systèmes experts. Ces items ont ainsi été choisis en fonction des critères habituels d'élaboration d'une base de données médicale, mais également en fonction de leur pertinence vis-à-vis du diagnostic (identification des germes).

Les faits initiaux que l'on peut fournir à DIAGSES et qui sont manipulés dans la base de connaissances pour déduire les probabilités bactériennes peuvent donc être générés à partir des variables de la base de données. Il en est de même pour ABSES qui utilise, pour fournir ses conseils thérapeutiques, les données bactériologiques et un sous-ensemble des signes manipulés par DIAGSES.

Les dossiers de la base de données peuvent donc être traduits en cas cliniques comparables à ceux habituellement soumis au système expert. L'ensemble de ces dossiers constitue une sélection nationale des cas de septicémie observés en France actuellement (chapitre III A), cas pour lesquels le diagnostic et le traitement sont connus. On dispose ainsi d'un échantillon test pour évaluer le système expert SE_SES.

LINXES est un outil d'aide à la validation et à la maintenance de SE_SES. LINXES exécute automatiquement DIAGSES ou ABSES sur l'ensemble ou une partie des dossiers de la base de données. Les principaux résultats qu'il fournit sont les suivants :

- a. DIAGSES : Pour chaque dossier, les données cliniques et biologiques décrivant l'état du patient, le diagnostic proposé par le système et le(s) germe(s) identifié(s) par l'examen bactériologique (diagnostic réel) sont archivés. De plus, le diagnostic réel constituant un critère objectif pour évaluer les performances de DIAGSES, LINXES fournit, pour chaque germe, le pourcentage de réponses correctes du système. On dispose ainsi d'un coefficient d'efficacité ou de fiabilité de SE_SES pour l'identification des germes.

- b. ABSES : Pour chaque dossier, les faits initiaux (données cliniques et bactériologiques), le traitement réel et les conseils thérapeutiques du système sont également archivés. Les experts disposent ainsi de l'information requise pour évaluer les performances thérapeutiques de SE_SES.

1.2 Structure de LINXES

LINXES a été réalisé de façon à prendre en compte les remarques suivantes :

- a. Le système expert n'utilise qu'un sous ensemble des variables de la base de données. Par exemple, la variable 'hématocrite' n'est associée à aucun fait manipulé par SE_SES.
- b. Les items manipulés par le système expert peuvent être déduits de ce sous ensemble de variables, mais une étape de traduction est nécessaire. Par exemple, si ab_ant=0 dans la base de données, il faut générer le fait SE_SES 'non antibiothérapie antérieure'.
- c. Il faut tenir compte du caractère évolutif de l'application : maintenance des bases de connaissances et de la base de données.
- d. Le logiciel doit pouvoir être utilisé dans le cadre d'une autre application médicale, c'est à dire avec une autre base de données et une autre base de connaissances pour l'aide au diagnostic ou au traitement.
- e. L'application médicale étant définie, la production des résultats de l'évaluation doit être automatique.

LINXES comporte ainsi trois modules indépendants écrits en langage C :

- Le premier module, **LINK1**, est spécifique à l'application **SES**. Il permet la création d'un fichier de synthèse regroupant toutes les informations nécessaires à la construction des cas cliniques reconnus par **SE_SES**. La création de ce fichier est justifiée par la structure 'éclatée' des données provenant de la base de données.
- Le deuxième module, **LINK2**, permet la création et la modification d'un dictionnaire dont la fonction est de filtrer les variables de la base de données et de traduire les variables retenues en faits reconnus par le système expert. Ce module permet de décrire l'application médicale.
- Le troisième module, **LINK3**, réalise automatiquement le traitement. Il utilise un fichier unique de données (pour l'application **SES**, le fichier de synthèse) et le dictionnaire précédent pour générer les cas cliniques, exécuter le système expert sur ces cas cliniques et fournir les résultats de ses performances.

1.3 Description des modules

1.3.1 **LINK1**

La structure de la base de données implantée sur les "Rainbow" des 19 centres participant au projet a été décrite précédemment. Les dossiers sont transférés anonymement sur le VAX 780 du CERIM au début de chaque année pour participer à l'ensemble des études du projet. Au cours de cette étape de centralisation, un prétraitement permet d'éliminer les dossiers inexploitable. On obtient ensuite sept

fichiers de données à l'image des chapitres du dossier :

- IDSES. Identification, résultats bactériologiques et suivi thérapeutique.
- HISSES. Histoire clinique.
- ECSES. Examen clinique.
- BIOSES. Données biologiques.
- LSSES. Localisations septiques.
- TGSSES. Terrain général.
- CDSSES. Conclusion du dossier.

Les sept fichiers précédents comportent plus de 600 items cliniques et biologiques sur les septicémies. Le système expert n'utilise qu'un sous-ensemble de ces variables contenues dans différents fichiers. D'autre part, la dispersion des données dans différents fichiers était incompatible avec une exécution rapide du système expert sur l'ensemble des dossiers archivés. Nous avons donc choisi de regrouper toutes les informations permettant de générer des cas cliniques, compréhensibles par SE_SES, dans un fichier de synthèse.

Le dessin de ce fichier de synthèse a été établi en collaboration avec les experts, en déterminant les variables de la base de données permettant de générer des faits SE_SES. Ces variables seront appelées dans la suite variables utiles. Par exemple, le fichier CDSSES ne contient aucune variable utile, le fichier IDSES comporte 27 variables, parmi lesquelles 15 sont des variables utiles.

LINK1 construit le fichier de synthèse à partir des tables suivantes :

- Une table par fichier retenu pour le transfert. Chacune de ces tables fournit la liste des variables utiles contenues dans le fichier correspondant.
- Une table globale fournit la liste des fichiers retenus et associe à chacun de ces fichiers sa table de variables utiles.

La FIGURE 1.I illustre le fonctionnement de LINK1.

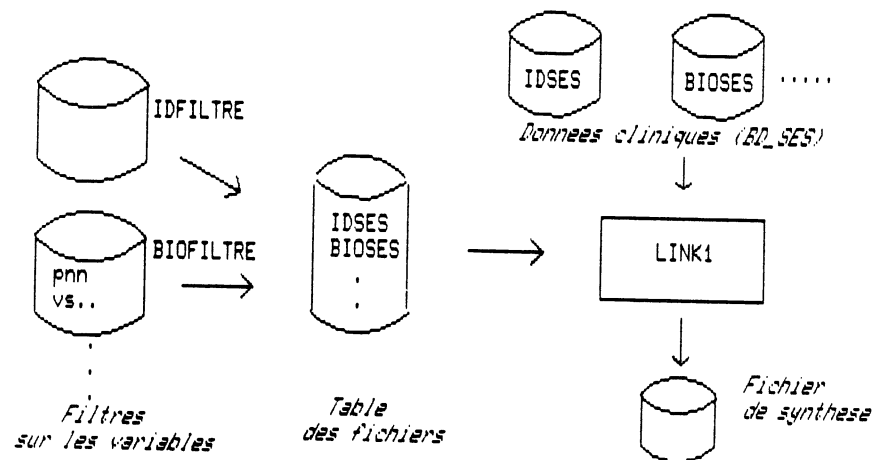


FIGURE 1.I

Le fichier de synthèse a été créé une fois pour toutes. Cependant, la structure de la base de données ou le dictionnaire des items manipulés SE_SES peuvent évoluer (maintenance des bases de connaissances). Dans les deux cas, il pourra être nécessaire de mettre à jour les tables précédentes et de réexécuter LINK1 pour obtenir le nouveau fichier de synthèse correspondant. Le fichier de synthèse comporte actuellement 85 variables.

1.3.2 LINK2 : gestion du dictionnaire de l'application

Les faits transmis au système expert pour chaque patient sont de deux types :

- a. les items binaires : "infection orl", "non antibiothérapie antérieure" ...
- b. les items valuables : "température=39.5" ...

Nous allons voir comment ces faits sont générés à partir des données du fichier de synthèse. La **FIGURE 1.II** montre la structure du fichier de synthèse.

		Patient 1	Patient i	Patient n
<i>Identification</i>	Sexe		M	
	Age		65	
<i>Histoire clinique</i>	Antibio_ant		0	
	Lieu_contr		2	
<i>Examen clinique</i>	Temperature		39.5	
	Choc		1	
<i>Biologie</i>	PNN		72	
	VS		31	
<i>Localisations septiques</i>	LS2		24	
	LS1		53	
<i>Terrain general</i>	TG1		101	
	TG2			

FIGURE 1.II

Il comporte des variables quantitatives (âge...) et des variables qualitatives (localisations septiques...). Pour les variables qualitatives, on connaît les valeurs des modalités : dans notre exemple, le patient i possède la modalité 24 (sinusite) pour la première variable localisation septique. Un item binaire SE_SES peut être généré à partir d'une modalité (si "terrain général"=101, on doit générer le fait "tumeur_intestin") ou de plusieurs modalités : aux modalités 24 (sinusite) ou 25 (otite) d'une variable "localisation septique" correspondent le fait "infection_orl". Certaines modalités ne sont pas reconnues par SE_SES, comme la modalité 92 (lymphangite). La FIGURE 1.III représente ainsi les liens possibles entre les items de la base de données et ceux manipulés par SE_SES.

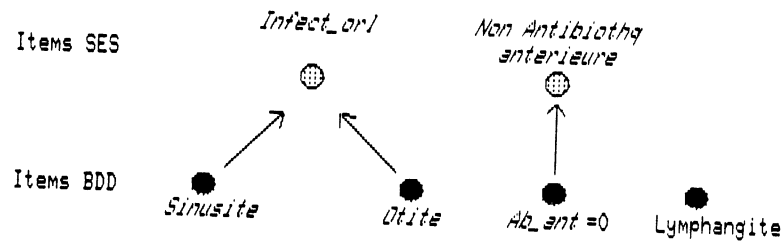


FIGURE 1.III

Certaines variables qualitatives comportent un nombre important de modalités (par exemple, les variables 'localisation septique' ont 61 modalités). Nous avons regroupé ces variables qualitatives en cinq classes sémantiques :

- a. germe(s) identifié(s) par l'examen bactériologique.
- b. antibiotiques prescrits.
- c. localisations septiques.
- d. terrain général.
- e. toutes les autres : lieu d'acquisition, porte d'entrée...

Une sixième classe est associée aux variables numériques, pour lesquelles nous devons réaliser une correspondance synonymique et transmettre au système expert leur valeur.

A chacune des classes précédentes, nous avons associé une table de correspondance permettant de traduire les données utiles du fichier de synthèse en faits reconnus et donc manipulables par le système expert. Ces tables de correspondance ont été établies en collaboration avec les experts.

Une table globale contient la liste des variables utiles et associe à chacune de ces variables sa table de correspondances. Cette table permet de filtrer les données à soumettre au système : seules seront retenues et traduites les variables présentes dans cette table.

Pour DIAGSES, la table globale contient l'ensemble des variables du fichier de synthèse, à l'exception des données bactériologiques et thérapeutiques. Pour ABSES, elle contient les données cliniques utiles et les données bactériologiques.

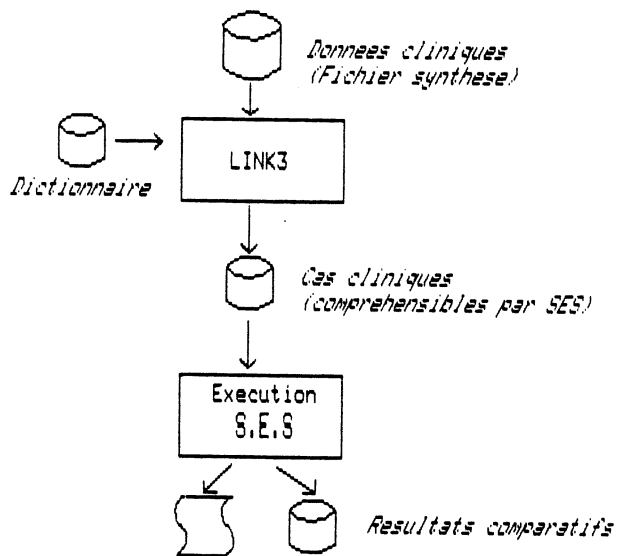
L'ensemble de ces tables constitue le dictionnaire de l'application SES, géré par le module LINK2. Un dictionnaire analogue peut être construit pour un autre système expert médical. LINK2 permet la création et la modification d'un tel dictionnaire. Celui-ci est construit une fois pour toutes, pour une application donnée. Il peut cependant être nécessaire de le mettre à jour en cas de modification dans la structure de la base de données ou en cas de modification des items manipulés par le système expert.

1.3.3 LINK3 : le module de transfert-résultats fournis

Le module de transfert (LINK3) est entièrement automatique. Il peut être exécuté sur l'ensemble des dossiers de la base de données ou sur un groupe de dossiers (par exemple ceux d'une année choisie). LINK3 utilise le dictionnaire précédent pour filtrer les

enregistrements du fichier de synthèse. Pour chaque patient, les items reconnus par SE_SES sont générés grâce au dictionnaire et constituent automatiquement la base de faits de DIAGSES ou ABSES. Simultanément, LINK3 extrait des enregistrements le traitement réel suivi par le patient ou le micro-organisme identifié par les résultats bactériologiques. SE_SES est ensuite exécuté sur la base de faits ainsi construite. LINK3 fournit enfin les résultats permettant l'évaluation du système expert (FIGURE 1.IV).

FIGURE IV



Détaillons les résultats fournis dans le cas de DIAGSES. Dans les conclusions de DIAGSES, on retrouve une liste de germes, généralement un, deux ou trois germes, classés par ordre décroissant de vraisemblance. Nous avons adopté la même échelle de valeurs que celle retenue par les experts lors des autres validations du système (chapitre III a) :

- note 1 (bonne réponse) lorsque le germe réel est identifié en première position.
- note 2 (bon avis) lorsque le germe réel est cité dans l'une des trois premières positions.
- note 3 lorsque le germe n'est pas cité ou n'est cité qu'après la troisième position.

Dans les conclusions de DIAGSES, certains germes peuvent apparaître avec des coefficients de vraisemblance égaux. Pour les notes 1 et 2 respectivement, nous avons considéré que la réponse du système était correcte, lorsque le germe réel était cité dans la liste des ex-aequo.

Outre la liste, pour chaque dossier, des faits initiaux, du diagnostic réel et des conclusions du système, les résultats fournis par LINK3 sont résumés de la manière suivante :

- a. Un tableau (Diagnostics_réels) X (note). Ce tableau permet d'obtenir, pour chaque germe, le pourcentage d'identification en première position (respectivement dans l'une des trois premières positions, et après la troisième position) (FIGURE 1.V). On obtient ainsi un coefficient d'efficacité du système pour l'identification des germes, sur les observations de la base de données.

CHAPITRE III B: - EVALUATION DU SE_SES -

RESULTATS BACTERIOLOGIQUES	NOTE 1	NOTE 2	% NON IDENTIFIES NOTE 3	
ECOLI	E1%	E2%	E3%	100%
PNEUMOCOQUE	P1%	P2%	P3%	100%
.				
.				

FIGURE 1.V

- b. Les dossiers pour lesquels la note est différente de 1 sont mis en évidence et archivés, ce qui rend possible leur analyse par les experts.
- c. Un tableau (Diagnostics_réels) X (Premières Conclusions de DIAGSES). Ce tableau permet d'étudier, pour chaque germe, la qualité de la discrimination du système (FIGURE 1.VI). On peut observer, pour chaque germe, la ventilation des réponses du système.

RESULTATS BACTERIOLOGIQUES	CONCLUSIONS DIAGSES	ECOLI	PNEUMOCOQUE	.	.	.
ECOLI		EE ¹	EP	.		
PNEUMOCOQUE		PE	PP	.		
.		.	.	.		
.		.	.	.		

¹EE : nombre d'"ECOLI" réels classés par le système en "ECOLI".

FIGURE 1.VI

Comme nous le verrons dans la suite de notre exposé, LINXES nous a permis de fournir aux experts une évaluation précise des performances de DIAGSES sur les dossiers suffisamment documentés de la base de données et de tester ensuite les modifications apportées à la base de connaissances.

Par ailleurs, ce logiciel constitue un outil d'aide à la maintenance de SE_SES. Chaque année, 450 dossiers supplémentaires provenant des différents centres participant au projet SES seront en effet archivés. Les résultats de l'évaluation des performances du système sur ces nouveaux dossiers seront fournis au groupe d'experts chargé de la maintenance des bases de connaissances.

2 EVALUATION DE LA BASE DE CONNAISSANCES ORIGINELLE (DIAGSES)

Nous avons utilisé LINXES pour évaluer les performances de la base de connaissances originelle de DIAGSES, en ce qui concerne l'aide à l'identification des germes. Cette évaluation a été réalisée sur l'ensemble des dossiers interprétables de la base de données.

Notre objectif était de répondre aux questions suivantes :

- Quels sont les germes correctement (ou difficilement) identifiés par le système ?
- Quelles sont les principales erreurs de discrimination du système ?

- Quelles améliorations peut-on apporter au système ?

Le fichier de synthèse comporte 410 septicémies pour l'année 1985 et 406 pour l'année 1986. L'étude a été réalisée séparément pour 1985 et 1986, de façon à pouvoir comparer les fréquences des germes et les résultats de l'évaluation sur ces 2 années.

2.1 Critères d'évaluation et de sélection des dossiers.

Avant toute évaluation, il est nécessaire de définir ce que l'on veut évaluer. Il peut, par exemple, être intéressant de mesurer la rapidité du système, sa facilité d'utilisation, ou la modularité de la base de connaissances (chapitre I). Nous nous intéresserons, pour notre part, à la justesse diagnostique du système. Nous savons que DIAGSES fournit dans ses conclusions les germes probablement responsables de l'infection, mais aussi des diagnostics intermédiaires et des conseils d'explorations complémentaires. Notre étude ne concerne que la partie identification des germes.

Il est également nécessaire de déterminer quelle est la réponse correcte et comment on peut l'obtenir [GASCHNIG 83]. Dans le cadre de notre application, DIAGSES, il s'agit du germe responsable de l'infection ; ce germe est connu dès que l'on dispose des résultats des examens bactériologiques. Nous comparerons donc les conclusions du système au germe réel, enregistré dans le dossier BD_SES.

Il est d'autre part nécessaire de définir un échantillon test, aussi large que possible [WYATT 87] : il s'agira ici des dossiers interprétables de la base de données des années 1985-1986.

Plus précisément, parmi les 816 dossiers correspondants du fichier de synthèse, nous avons écarté de l'étude :

- a. les dossiers pour lesquels les résultats bactériologiques étaient inconnus : 29 en 1985 et 22 en 1986.
- b. les dossiers pour lesquels les résultats bactériologiques étaient imprécis (bacille gram négatif sans autre précision par exemple) : 4 en 1985 et 5 en 1986.
- c. les patients présentant une infection polybactérienne : 29 dossiers en 1985 et 24 en 1986. Ceux-ci pourraient faire l'objet d'une étude spécifique : il semble intéressant de savoir si le système reconnaîtrait au moins l'un, sinon davantage, des germes incriminés.
- d. les dossiers insuffisamment documentés :

Rappelons que les faits que l'on peut introduire dans DIAGSES appartiennent à quatre familles principales : le terrain général (antécédents,...), la localisation septique (urogénitale,...), l'allure clinique de l'infection (lieu d'acquisition,...) et des données biologiques (numération, formule sanguine,...). A partir d'éléments appartenant à ces différentes familles, DIAGSES déduit les probabilités bactériennes. Pour conclure, le système doit cependant disposer d'une information suffisante. Il existe des observations réelles de septicémie, pour lesquelles l'information disponible ne permet généralement pas aux experts de poser le diagnostic : par exemple, les septicémies à choc toxi-infectieux, en l'absence de tout autre syndrome.

Nous avons contrôlé l'existence, dans le dossier, des faits qui doivent être obligatoirement connus : sexe, âge, température. D'autre part, avec les experts, nous avons choisi d'éliminer les dossiers pour lesquels le lieu d'acquisition de l'infection (nosocomial ou communautaire) était inconnu, et ceux pour lesquels aucune localisation septique, ni simultanément aucun élément du terrain général, n'avaient été mis en évidence. 70 dossiers ont ainsi été éliminés pour l'année 1985 et 78 pour l'année 1986.

L'étude a donc porté sur 555 dossiers.

2.2 Performances globales et par germe

Rappelons que la base de connaissances de DIAGSES se propose de différencier 25 germes retenus comme possibles par les experts. LINXES permet d'obtenir tout d'abord les fréquences d'apparition de ces germes dans notre population (TABLEAU2.I).

GERMES	1985		1986	
	EFFECTIFS	FREQUENCES	EFFECTIFS	FREQUENCES
ACINETOBACTER	0	0	1	0,36%
BACTEROIDE	0	0	4	1,44%
BRUCELLE	2	0,72%	0	0
CANDIDA	3	1,08%	1	0,36%
CLOSTRIDIUM	1	0,35%	1	0,36%
ENTEROBACTER	3	1,08%	4	1,44%
ENTEROCOQUE	10	3,6%	17	6,14%
ECOLI	68	24,46%	66	23,83%
GONOCOQUE	0	0	1	0,36%
HEMOPHILUS	3	1,08%	3	1,08%
KLEBSIELLE	8	2,87%	10	3,61%
LISTERIA	3	1,08%	6	2,17%
MENINGOCOQUE	10	3,6%	4	1,44%
PNEUMOCOQUE	42	15,11%	32	11,55%
PROTEUS	9	3,24%	8	2,89%
PSEUDOMONAS	5	1,8%	7	2,53%
SALMONELLE	17	6,11%	9	3,25%
SERRATIA	0	0	3	1,08%
STAPH DORE	54	19,43%	56	20,23%
STAPH BLANC	6	2,16%	9	3,25%
STREPTO A	6	2,16%	7	2,53%
STREPTO B	6	2,16%	4	1,44%
STREPTO D	9	3,24%	9	3,25%
STREPTO NG	12	4,32%	11	3,97%
YERSINIA	0	0	1	0,36%
AUTRES GERMES	1	0,35%	3	1,08%

TABLEAU 2.I

Ces résultats épidémiologiques sont comparables à ceux obtenus par nos études antérieures [PORTIER 87]. Ils montrent, par ailleurs, que les 2 populations 1985 et 1986 sont comparables : il n'existe pas de différences significatives (au seuil de 5%) entre les fréquences observées pour les différents germes selon ces 2 années. On remarque un nombre important de germes très peu fréquents (moins de 5 cas observés). Les germes les plus fréquents sont l'Escherichia Coli (24,3%), le Staphylocoque Doré (20%) et le Pneumocoque (13,4%).

CHAPITRE III B: - EVALUATION DU SE_SES -

LINXES fournit aussi un tableau (GERMES_REELS)x(NOTES). Rappelons que la note 1 (respectivement 2) est attribuée à un dossier lorsque le germe responsable de l'infection est identifié en première position (respectivement dans l'une des trois premières positions). La note 3 est attribuée lorsque le germe n'est pas cité par le système dans l'une des 3 premières positions.

Le pourcentage global de germes identifiés en première position par DIAGSES (note 1) est de 50% pour 1985 et de 47% pour 1986. DIAGSES donne un "bon avis" (note 2) dans 68% des cas en 1985 et dans 64% en 1986. Ces pourcentages globaux d'identification ne sont pas statistiquement différents (au seuil de 5%). Le TABLEAU 2.II montre quels sont les germes les mieux identifiés par le système.

GERMES	1985			1986		
	EFFECTIFS	NOTE 1	NOTE 2	EFFECTIFS	NOTE 1	NOTE
ECOLI	68	88,24%	95,6%	66	83,3%	95,1%
PNEUMOCOQUE	42	78,57%	90,5%	32	68,75%	72%
MENINGOCOQUE	10	70%	80%	4	75%	100%
STAPHYLOCOQUE DORE	54	48,15%	74%	56	67,9%	84%
STREPTO_A	6	33,4%	66,7%	7	29%	86%
STREPTO_NON GROUPABLE	12	58,3%	66,7%	11	27%	46%
LISTERIA	3	0	66,7%	6	0	100%

TABLEAU 2.II

En 1985 et 1986, ce sont les mêmes germes qui sont bien identifiés. Les pourcentages de bonnes réponses du système pour ces germes sont comparables selon ces 2 années : il n'existe pas de différence significative entre les fréquences d'identification observées (au seuil de 5%).

CHAPITRE III B: - EVALUATION DU SE_SES -

Les germes les moins bien reconnus par le système sont donnés par le TABLEAU 2.III.

GERMES RARES	1985		1986	
	EFFECTIFS	BON AVIS	EFFECTIFS	BON AVIS
ACINETOBACTER	0	x	1	0
BACTEROIDE	0	x	4	1
CANDIDA	3	0	1	0
CLOSTRIDIUM	1	0	1	0
GONOCOQUE	0	x	1	0
SERRATIA	0	x	3	0
YERSINIA	0	x	1	0
AUTRES GERMES				
ENTEROBACTER	3	0	4	1
ENTEROCOQUE	10	1	17	0
KLEBSIELLE	8	2	10	3
PROTEUS	9	4	8	2
PSEUDOMONAS	5	1	7	2
STAPHYLOCOQUE BLANC	6	1	9	3
STREPTO_B	6	0	4	0

TABLEAU 2.III

Les tableaux II et III résument les performances du système en terme de sensibilité.

Les germes à forte prévalence, l'Escherichia Coli, le Staphylocoque Doré et le Pneumocoque, sont bien identifiés par le système.

Les performances du système concernant les germes très rarement observés dans notre population (moins de 5 cas observés sur les 555 septicémies) sont données à titre indicatif. Elles devront être vérifiées sur un échantillon plus large.

Etudions les résultats suivant les classes bactériologiques :

- a. Les **COCCI GRAM POSITIF** : le Pneumocoque, le Streptocoque A, le Streptocoque Non Groupable et le Staphylocoque Doré sont bien identifiés. Les Streptocoques Bovis et Entérocoque, ainsi que le Staphylocoque Blanc, sont plus difficilement reconnus.
- b. Les **COCCI GRAM NEGATIF** : le Méningocoque est très bien reconnu par DIAGSES. Le Gonocoque est quasiment absent de notre population.
- c. Les **BACILLES GRAM POSITIF** : il s'agit pour SE_SES du Listeria. ce germe est bien identifié.
- d. Les **BACILLES GRAM NEGATIF** : l'Escherichia Coli bénéficie d'un très bon pourcentage d'identification. L'Hémophilus est également assez bien reconnu par le système. La Salmonelle est assez bien identifiée, mais en deuxième ou troisième position. Le Proteus, la Klebsielle et le Pseudomonas sont plus difficilement identifiés. Nous étudierons particulièrement les réponses du système concernant ces germes dans les prochains paragraphes.
- e. Les **ANAEROBIES** et les **CANDIDA** : ces germes sont très peu fréquents dans notre population. Les pourcentages d'identification correspondants ne sont donc pas interprétables.

Cette étude nous a fourni un bilan quantitatif des performances du système pour l'identification de chaque germe. Le paragraphe suivant sera consacré à l'étude de la qualité de la discrimination.

2.3 Analyse des réponses du système

Le module de transfert fournit également le tableau de fréquences (GERMES_REELS) x (PREMIERE_CONCLUSION) permettant d'étudier la qualité de la discrimination des germes par le système (TABLEAU 2.IV).

GERME REEL \ GERME DIAGSES	ECOLI	PNEUMOCOQUE	STAPH_DORE	...	TOTAL (GERMES REELS)
ECOLI	EE	EP	ES	...	ECOLI
PNEUMOCOQUE	PE	PP	PS	...	PNEUMOCOQUE
STAPH_DORE	SE	SP	SS	...	STAPH_DORE
.	.				
.	.				
TOTAL (GERMES DIAGSES)	ECOLI	PNEUMOCOQUE	STAPH_DORE	...	555

TABLEAU 2.IV

Dans ce tableau, on retrouve le pourcentage d'identification en première position pour chaque germe (note 1). Mais il permet d'analyser plus précisément les performances du système en observant, pour chaque germe, la nature des réponses de DIAGSES et les erreurs de classement. Ce tableau fait apparaître en particulier, pour chaque micro-organisme, le nombre de faux positifs et de faux négatifs. Les résultats étant semblables pour les 2 années, nous n'étudierons que les données globales.

Le TABLEAU 2.V montre la qualité de la discrimination du système pour la première conclusion en ce qui concerne les 3 germes les plus fréquents.

CHAPITRE III B: - EVALUATION DU SE_SES -

RESULTATS BACTERIOLOGIQUES	CLASSEMENT DIAGSES		TOTAL GERMES REELS
	Staph_Doré	non Staph_Doré	
Staph_Doré	64	46	110
non Staph_Doré	63	378	441
TOTAL DIAGSES	127	424	555
	Pneumocoque	non Pneumocoque	
Pneumocoque	52	22	74
non Pneumocoque	26	451	477
TOTAL DIAGSES	78	473	555
	Ecoli	non Ecoli	
Ecoli	115	19	134
non Ecoli	95	322	417
TOTAL DIAGSES	210	341	555

TABLEAU 2.V

Les fréquences du Staphylocoque Doré et du Pneumocoque ne sont pas surestimées par le système. L'étude du tableau (GERMES_REELS) x (PREMIERE_CONCLUSION) fait apparaître, pour les trois germes du tableau 2.V, les résultats suivants :

- 64 Staphylocoques Dorés sont reconnus par DIAGSES dès la première conclusion. 46 Staphylocoques Dorés ne sont pas identifiés en première position. Cependant, pour 14 d'entre eux, le système a bien indiqué un Staphylocoque mais n'a pas précisé s'il s'agissait d'un Staphylocoque Doré ou d'un Staphylocoque Blanc. 11 Staphylocoques Dorés ont été classés en Escherichia Coli, 7 en Streptocoque Non Groupable et 6 en Pneumocoque.

- 63 Staphylocoques Dorés ont été proposés à tort en première position par le système. Parmi ceux-ci, 10 étaient des Escherichia Coli, 9 des Streptocoques-Entérocoque et 23 des autres Streptocoques.
- 52 Pneumocoques sont reconnus en première position par DIAGSES. Parmi les 22 Pneumocoques non identifiés, 5 ont été classés en Méningocoque, 4 en Escherichia Coli et, pour 7 cas, aucune réponse n'a été proposée.
- Les 26 Pneumocoques proposés à tort par le système en première position étaient principalement des Staphylocoques (6), des Streptocoques (6) et des Listeria (5).
- L'Escherichia Coli est très bien identifié par le système (86%). Cependant, sa fréquence est surestimée par DIAGSES. Ce germe est en effet cité 210 fois en premier par le système, pour un effectif réel de 134.

Nous avons donc examiné particulièrement ce problème. Ainsi, le **TABLEAU 2.VI** montre quels sont les germes les plus fréquemment reconnus à tort comme un Escherichia Coli.

GERMES RARES	EFFECTIFS	BONNES REponses	ECOLI	AUTRES GERMES	PAS DE CONCLUSION
ACINETOBACTER	1	0	1	0	0
BACTEROIDE	4	0	3	1	0
CANDIDA	4	0	2	1	1
CLOSTRIDIUM	2	0	2	0	0
GONOCOQUE	1	0	1	0	0
SERRATIA	3	0	2	1	0
YERSINIA	1	0	1	0	0
<hr/>					
AUTRES GERMES					
ENTREROBACTER	7	0	3	4	0
KLEBSIELLE	18	1	10	6	1
PROTEUS	17	0	8	7	2
PSEUDOMONAS	12	2	6	3	1
SALMONELLE	26	0	20	6	0

TABLEAU 6.VI

Ce tableau identifie 59 "faux" Escherichia Coli sur un total de 76.

On y retrouve les germes très rarement observés dans notre population. Il semble donc que ces germes soient reconnus en première position comme des Escherichia Coli, mais ce résultat ne constitue qu'une indication, compte tenu de leur très faible fréquence.

Les autres germes figurant dans ce tableau sont, à l'exception du Streptocoque B, de l'Entérocoque et du Staphylocoque Blanc, les mêmes que ceux que DIAGSES reconnaît difficilement.

Lorsqu'ils ne sont pas identifiés en première position, ces germes sont donc principalement mal classés en Escherichia Coli. Cette constatation est particulièrement évidente pour la Salmonelle. Nous verrons cependant, dans notre conclusion, que cette situation s'explique en partie par la ressemblance entre les tableaux cliniques

de certains de ces germes et de l'Escherichia Coli.

Le Staphylocoque Blanc, l'Entérocoque et le Streptocoque B sont les seuls germes difficilement identifiés par le système qui ne figurent pas dans ce tableau. Le Staphylocoque Blanc est principalement classé par le système en Staphylocoque Doré, plus rarement en Klebsielle et en Escherichia Coli. L'Entérocoque est classé, dans exactement les mêmes proportions, en Escherichia Coli, Staphylocoque Doré et Streptocoque Non Groupable. Enfin, le Streptocoque B est proposé principalement en Streptocoque sans autre précision ou en Pneumocoque.

2.4 Conclusion de l'évaluation

Nous avons analysé l'ensemble de ces résultats avec les experts du GROUPE SES. Cette analyse nous a permis de situer notre évaluation par rapport aux autres évaluations réalisées sur DIAGSES et de déterminer, compte tenu du niveau actuel de l'expertise, les améliorations que l'on pouvait apporter au système.

2.4.1 **Comparaison avec les évaluations antérieures.**

Rappelons que DIAGSES a bénéficié par ailleurs de trois évaluations complémentaires fondées sur trois critères [CBEUSCART 88] :

- Mesure de l'accord entre DIAGSES et le diagnostic réel. Les pourcentages de concordance et de discordance entre les conclusions du système et le germe responsable de l'infection, lorsqu'il est mis en évidence par l'examen bactériologique, ont été évalués sur 115 cas réels de septicémie provenant des

différents centres participant au projet.

- La confrontation entre DIAGSES et différentes catégories d'acteurs. Il s'agissait d'établir si les performances du système pouvaient être distinguées de celles d'étudiants de fin d'étude, de médecins généralistes, de médecins suivant une formation complémentaire en infectiologie et de spécialistes en infectiologie. Nous avons ainsi situé les performances de DIAGSES au sein de cette hiérarchie de compétences.
- L'évaluation prospective dans les différents centres du GROUPE SES. Cette évaluation, poursuivie actuellement, a pour but de confronter la réalité clinique à la réponse de DIAGSES, de détecter d'éventuelles erreurs ou incohérences, et d'améliorer l'utilisation du système.

Notre évaluation rentre dans le cadre de la première évaluation. Cependant, la méthodologie employée (l'exécution globale et automatique sur tous les dossiers de la base de données et la nature des résultats fournis), amène les remarques suivantes :

- a. Nous ne mettons pas en évidence les réponses fausses : par exemple, une septicémie avec infection urinaire pour laquelle DIAGSES proposerait une Salmonelle.
- b. Quantité et qualité de l'information soumise au système expert.

Pour fournir des conseils pertinents, tout système d'aide à la décision doit disposer d'une information suffisante et fiable.

La base de données est implantée dans les différents centres du GROUPE SES avec possibilité d'exploitation locale. Il y a participation active de ces centres à l'ensemble du projet. Toutes les garanties sont ainsi prises pour assurer un remplissage correct des dossiers.

Il existe cependant des cas réels de septicémie pour lesquels la quantité d'information connue est insuffisante, au niveau actuel de l'expertise, pour prédire le germe en cause : par exemple, une septicémie à choc-toxi infectieux, en l'absence de tout autre syndrome.

En accord avec les experts, nous avons éliminé les dossiers pour lesquels, ni le terrain général, ni la localisation septique, n'étaient renseignés. Il s'agit d'un minimum en-dessous duquel les experts eux-mêmes ne disposent généralement pas d'informations suffisantes pour poser un diagnostic.

Au cours de la première évaluation, les 115 patients étudiés étaient des dossiers réels représentatifs de l'ensemble des septicémies, mais ces dossiers étaient tous bien renseignés. Nous ne pouvons obtenir, par une sélection automatique des dossiers, une garantie équivalente.

D'autre part, lors de l'évaluation portant sur ces 115 cas, les dossiers ont été rentrés par des médecins. Nous étions ainsi assurés de la qualité des informations fournies au système expert.

- c. Avec notre méthode, nous obtenons un coefficient d'efficacité absolue du système, puisque la référence est le diagnostic réel. Une évaluation complète doit également prendre en compte le niveau

actuel de l'expertise, comme c'était le cas pour la deuxième évaluation.

Cette dernière remarque nous conduit à étudier les possibilités et impossibilités de discrimination.

2.4.2 Germes difficilement identifiables.

Nos résultats confirment que, dans le cadre de notre application, compte tenu du niveau actuel de l'expertise, il existe des quasi-impossibilités de discrimination.

Elles concernent tout d'abord des germes qui ont une très faible prévalence et qui correspondent à des cas très particuliers. Par exemple, sauf chez une femme après avortement septique, le Clostridium ne sera pas proposé en première intention par les experts. Ces cas ne correspondent pas au profil général de nos services de maladies infectieuses. C'est une explication de la difficulté qu'a DIAGSES à identifier les germes à très faible prévalence. Relevons que l'analyse des données ne peut pas actuellement apporter une aide à l'identification de ces germes, compte tenu de leurs fréquences dans notre échantillon.

Ces difficultés de discrimination concernent aussi des germes dont les tableaux cliniques sont très voisins. Ainsi, pour les experts, la Klebsielle à porte d'entrée urinaire ressemble à l'Escherichia Coli, tandis que la Klebsielle à porte d'entrée pulmonaire est proche du Pneumocoque. Les tableaux cliniques de l'Escherichia Coli et du Proteus à porte d'entrée urinaire leur apparaissent également très voisins. Il en est de même pour

l'Entérocoque à porte d'entrée urinaire, toujours avec l'Escherichia Coli. On trouve ainsi une explication à la prépondérance de l'Escherichia Coli dans les conclusions du système : 10 Klebsielles, 8 Proteus et 9 Entérocoques sont classés en Escherichia Coli.

2.4.3 Possibilités d'améliorations

Les résultats de notre évaluation doivent donc être interprétés en tenant compte de l'ensemble des remarques précédentes. Nous obtenons, par notre méthode, un coefficient d'efficacité ou de fiabilité du système pour l'ensemble des germes et sur les dossiers de la base de données sélectionnés suivant notre critère. Ce coefficient constitue un élément d'information parmi les autres évaluations. De ce point de vue, le pourcentage d'identification des germes en première position retrouvé par notre étude (près de 50%) est comparable à celui obtenu lors de l'évaluation effectuée sur les 115 dossiers bien renseignés. Ce coefficient d'efficacité constitue par ailleurs une référence permettant de juger de la qualité des modifications apportées à la base de connaissances de DIAGSES (dans le cadre de notre étude ou de la maintenance du système).

Notre étude montre cependant que les performances du système peuvent être améliorées. En particulier, l'identification, dès la première conclusion, des différents Staphylocoques et Streptocoques, de la Salmonelle et du Pseudomonas peut être améliorée.

D'autre part, la surrestimation de la fréquence de l'Escherichia Coli dans les premières conclusions du système doit être corrigée. Il faut cependant distinguer, pour ce dernier point, deux problèmes de difficulté différente :

- a. Il existe, comme nous l'avons relevé, des erreurs de discrimination qui étaient prévisibles (Ecoli - Klebsielle, Ecoli - Proteus, Ecoli - Entérocoque) et qui sont probablement difficilement corrigeables.
- b. D'autres erreurs doivent, au contraire, être corrigées : pour les experts, les confusions Ecoli - Pseudomonas, Ecoli - Salmonelle ne se justifient pas intellectuellement : ces germes ne correspondent pas à des tableaux cliniques voisins.

A partir de ces résultats, nous avons cherché quelles améliorations nous pouvions apporter à la base de règles utilisée par DIAGSES.

Nous avons tout d'abord déterminé s'il était possible, à partir des données du fichier informatisé, de différencier les germes ayant, pour les experts, des tableaux cliniques voisins.

3 LES DIFFICULTES DE DISCRIMINATION

Dans ce paragraphe, nous étudions deux problèmes considérés a priori comme difficiles par les experts :

- Le diagnostic différentiel entre les septicémies causées par l'Escherichia Coli, le Proteus ou la Klebsielle.
- L'identification des germes à faible prévalence dans les septicémies.

3.1 Diagnostic différentiel "Ecoli" "Proteus" "Klebsielle"

L'objectif principal de DIAGSES est l'identification du germe le plus probablement à l'origine de la septicémie, dans l'attente des résultats bactériologiques. Ce diagnostic repose essentiellement sur les données de l'examen clinique : le lieu d'acquisition de l'infection, les localisations septiques, les signes cliniques de l'infection, le terrain général et les antécédents du patient (chapitre III A). Cette démarche est équivalente à la démarche intellectuelle utilisée par les experts en maladies infectieuses.

Les données cliniques observées chez les patients qui souffrent d'une septicémie causée par l'Escherichia Coli ou le Proteus ou la Klebsielle sont très voisines.

Au niveau actuel de l'expertise, ces septicémies apparaissent plutôt comme un groupe homogène (groupe E-K-P) : il est possible d'identifier le groupe E-K-P, mais la distinction, au sein de ce groupe, entre les septicémies causées par l'Escherichia Coli, celles causées par le Proteus et celles causées par la Klebsielle, est très difficile.

Par exemple, lorsque la porte d'entrée du germe est urinaire, on sait que la septicémie est le plus souvent due à l'une de ces trois bactéries (dans 90% des cas de notre sélection : chapitre III A TABLEAU 3.V). Cependant, les experts ne disposent pas, actuellement, de données cliniques ou biologiques pertinentes pour déterminer quel est le germe causal. De plus, la fréquence de l'Escherichia Coli étant largement prépondérante par rapport à celles du Proteus ou de la Klebsielle (80%-10%-10%, pour cet exemple, dans notre sélection), la suspicion du clinicien, pour ces patients, se porte d'abord vers

l'Escherichia Coli.

Ces difficultés sont naturellement retrouvées par les résultats de notre évaluation : l'Escherichia Coli est très bien identifié par DIAGSES, alors que le Proteus et la Klebsielle sont difficilement reconnus et très souvent classés en première position, de manière erronée, en Escherichia Coli (TABLEAU 2.VI).

Nous avons étudié s'il était possible de déterminer, par les techniques de l'analyse des données, de nouvelles associations de signes, et donc de nouvelles règles, permettant de différencier l'Escherichia Coli, le Proteus et la Klebsielle.

Nous disposions, dans cet objectif :

- D'une base de cas cliniques, les septicémies de la base de données BD_SES causées par l'Escherichia Coli, le Proteus et la Klebsielle.
- De l'expertise des cliniciens du GROUPE SES : les connaissances de DIAGSES pour l'identification de ces trois micro-organismes.

Cette étude a été réalisée à partir de l'expertise existante, que nous allons tout d'abord examiner.

3.1.1 Etude de l'expertise

Les connaissances dynamiques de DIAGSES sont exprimées dans le formalisme des règles de production, pondérées par des coefficients de vraisemblance (CV) qui prennent en compte l'incertitude du raisonnement (chapitre III A) :

SI Prémisse ALORS Conclusion (CV).

Si la prémisse P est vraie, cette règle déduit une nouvelle connaissance, la conclusion C, avec la "vraisemblance" CV.

Nous avons étudié les règles de DIAGSES qui permettent de déduire "Escherichia Coli" (Ecoli), "Proteus" ou "Klebsielle". Les prémisses de ces règles sont des données de l'examen clinique, qui peuvent intervenir de manière isolée ou en association. Les conclusions, l'un des trois germes étudiés. La règle :

SI tumeur solide et chimiothérapie ALORS Ecoli (.5).
exprime la connaissance experte "la présence d'une tumeur solide associée à une chimiothérapie, est en faveur de l'Escherichia Coli avec la vraisemblance 0,5". La base de connaissances de DIAGSES comporte 47 règles qui déduisent "Ecoli", 20 "Proteus" et 36 "Klebsielle". Dans la suite de ce paragraphe, nous désignerons par signe une prémisse de l'une de ces règles qui peut donc être une donnée clinique ou une association de données cliniques.

L'étude de ces règles amène les remarques suivantes :

- a. Il existe des signes spécifiques au groupe E-K-P : "infection urinaire", "porte d'entrée urinaire", "lithiase rénale" ...
- b. Tout signe en faveur du Proteus avec une vraisemblance CV, est également en faveur de l'Escherichia Coli avec une vraisemblance supérieure. Nous en donnons ici trois exemples :

SI DIABETE	ALORS	PROTEUS (.2)
SI DIABETE	ALORS	ECOLI (.4)
SI INFECTION URINAIRE	ALORS	PROTEUS (.6)
SI INFECTION URINAIRE	ALORS	ECOLI (.9)
SI ACQUISITION MEDECINE ¹	ALORS	PROTEUS (.2)
SI ACQUISITION MEDECINE	ALORS	ECOLI (.5)

¹ : Le lieu d'acquisition est un service de médecine.

TABLEAU 3.I

c. Les signes en faveur de la Klebsielle sont aussi en faveur de l'Escherichia Coli avec une vraisemblance plus élevée. On relève cependant quatre exceptions :

1)	SI BPCO ¹	ALORS	KLEBSIELLE (.3)
	SI BPCO ET AB ANT ²	ALORS	STAPHYLOCOQUE (.3)
	SI BPCO ET NON AB ANT	ALORS	PNEUMOCOQUE (.4)
2)	SI B_PNEUMOPATHIE ³	ALORS	KLEBSIELLE (0.3)
	SI B_PNEUMOPATHIE	ALORS	PNEUMOCOQUE (0.5)
3)	SI B_PNEUMOPATHIE ET BPCO	ALORS	KLEBSIELLE (0.3)
	SI B_PNEUMOPATHIE ET BPCO	ALORS	PNEUMOCOQUE (0.5)
4)	SI B_PNEUMOPATHIE ET AB ANT	ALORS	KLEBSIELLE (0.3)
	SI B_PNEUMOPATHIE ET AB ANT	ALORS	STAPHYLOCOQUE (0.3)

¹ : BRONCHO-PNEUMOPATHIE CHRONIQUE OBSTRUCTIVE

² : ANTIBIOTHERAPIE ANTERIEURE.

³ : BRONCHO-PNEUMOPATHIE

TABLEAU 3.II

Les septicémies à infection pulmonaire peuvent donc être causées par la Klebsielle. On observe cependant que ces signes ne sont pas spécifiques à la Klebsielle, puisqu'ils sont également en faveur du Pneumocoque ou du Staphylocoque avec des vraisemblances supérieures (ou égales).

Individuellement, les signes (prémises des règles) a priori considérés comme pertinents pour identifier respectivement l'Escherichia Coli, le Proteus et la Klebsielle, n'ont pas été jugés discriminants par les experts, pour la distinction entre ces trois micro-organismes. Ce jugement se traduit dans le poids préférentiel attribué aux règles qui concluent à l'Escherichia. De plus, si l'on considère l'effet associatif de ces signes, le mode de combinaison des coefficients de vraisemblance (chapitre III A) favorise le diagnostic de l'Escherichia Coli, comme le montre l'exemple simple suivant d'un patient réel dont le diagnostic véritable était un Proteus :

SIGNES CLINIQUES	CONCLUSIONS DES REGLES ASSOCIEES (CV)		
	ECOLI	PROTEUS	KLEBSIELLE
ACQUISITION MEDECINE	0,5	0,2	
PORTE ENTREE URINAIRE	0,9	0,4	0,3
DIABETE	0,4	0,2	
CHOC	0,7	0,7	0,7

CONCLUSIONS DE DIAGSES:

ECOLI : 0,99
 PROTEUS : 0,88
 KLEBSIELLE: 0,79

Nous avons donc étudié s'il était possible de déterminer de nouvelles règles, utilisant des signes pertinents pour l'identification du groupe E-K-P, et qui permettraient de mieux différencier le Proteus et la Klebsielle de l'Escherichia Coli.

3.1.2 Méthode et Résultats

La base de cas cliniques comporte 169 patients, sélectionnés à partir des dossiers 1985-1986 de la base de données, selon les mêmes critères que pour l'évaluation de DIAGSES : 17 d'entre eux présentent un Proteus, 18 un Klebsielle et 134 un Escherichia Coli.

Les 48 variables retenues pour cette étude sont les variables cliniques utilisées, isolément ou en association, dans les règles qui concluent à "Ecoli", "Proteus" ou "Klebsielle" (règles du groupe E-K-P). La liste de ces variables est donnée dans le TABLEAU 3.III. Toutes ces variables sont binaires, à l'exception de l'âge et du nombre de polynucléaires neutrophyles (PNN).

VARIABLES		VARIABLES	
ABCES POUMON	(1=OUI/O=NON)	INFECTION URINAIRE	(1/0)
AGE		INSUFFISANCE RENALE	(1/0)
ANTIBIOTHERAPIE		LESION COLORECTALE	(1/0)
ANTERIEURE	(1/0)	LITHIASE BILIAIRE	(1/0)
ARTHRITE	(1/0)	LLC ¹	(1/0)
ASCITE	(1/0)	LYMPHOME	(1/0)
BPCO	(1/0)	MENINGITE	(1/0)
B PNEUMOPATHIE	(1/0)	NEUTROPENIE	(1/0)
BRÛLURE	(1/0)	OSTEITE	(1/0)
CHOC	(1/0)	PNN	
CIRRHOSE	(1/0)	PERITONITE	(1/0)
CHIMIOTHERAPIE	(1/0)	PORTE ENTREE CATHETER	(1/0)
ACQUISITION CHIRURGIE		PORTE ENTREE COLIQUE	(1/0)
DIGESTIVE	(1/0)	PORTE ENTREE URINAIRE	(1/0)
ACQUISITION CHIRURGICALE	(1/0)	PROTHESE OS	(1/0)
ACQUISITION HOSPITALIERE	(1/0)	PROSTATITE	(1/0)
ACQUISITION HEMATOLOGIE	(1/0)	PROTHESE VASCULAIRE	(1/0)
ACQUISITION REANIMATION	(1/0)	SONDE URINAIRE	(1/0)
ACQUISITION UROLOGIE	(1/0)	SPONDYLODISCITE	(1/0)
DERIVATION DU LCR	(1/0)	SUPPURATION BRONCHIQUE	(1/0)
DIABETE	(1/0)	THROMBOPHLEBITE SUPPUREE	(1/0)
ETHYLISME	(1/0)	TRAUMATISME CRANIEN	(1/0)
GROSSESSE	(1/0)	TUMEUR SOLIDE	(1/0)
HEMODIALYSE	(1/0)	TUMEUR INTESTINALE	(1/0)
INFECTION PELVIENNE	(1/0)	UROPATHIE	(1/0)

¹: Leucémie lymphoïde chronique.

TABLEAU 3.III

Analyse univariée.

Nous avons tout d'abord étudié le comportement individuel de chaque prémisse P des règles du groupe E-K-P vis à vis du diagnostic D (Ecoli, Klebsielle ou Proteus). P est donc soit une variable binaire, "porte entrée urinaire", soit une association de variables binaires,

"tumeur solide et chimiothérapie".

les tableaux de fréquences PxD ont montré que seules les variables "Porte d'entrée urinaire", "infection urinaire" et "antibiothérapie antérieure", étaient significativement liées au type de germe en cause ($p < 1/1000$). Sept signes, absents dans notre population E-K-P, ont alors été éliminés de l'étude : "abcès du poumon", "hémodialyse", "arthrite", "prothèse vasculaire", "acquisition en service d'hématologie", "leucémie lymphoïde chronique", et "infection pelvienne".

D'autre part, nous n'avons pas mis en évidence de différences significatives (au seuil 5%) entre les moyennes des variables quantitatives (âge et PNN) selon le type de germe. Individuellement, les variables sont donc peu prédictives.

Analyse discriminante

Nous avons ensuite réalisé une Analyse Factorielle Discriminante (AFD) sur les 41 variables renseignées de notre population.

Nous cherchons à différencier les trois groupes, "Escherichia Coli", "Proteus" et "Klebsielle". L'analyse fournit donc deux variables (ou composantes) discriminantes. Ces variables discriminantes sont les combinaisons linéaires des variables initiales dont les valeurs sont les plus différentes possibles pour les trois groupes étudiés. Elles peuvent permettre de déterminer de nouvelles associations de signes pertinentes pour le diagnostic, et donc de construire de nouvelles règles.

La **FIGURE 3.I** représente la projection des patients sur l'unique plan discriminant. Ce plan passe par les centres de gravité des trois groupes. Ce graphique permet de visualiser la qualité de la séparation des groupes par les composantes discriminantes. Elle sera d'autant meilleure que les centres de gravité des trois groupes seront éloigné et que les individus de chaque groupe seront proches de leur centre de gravité.

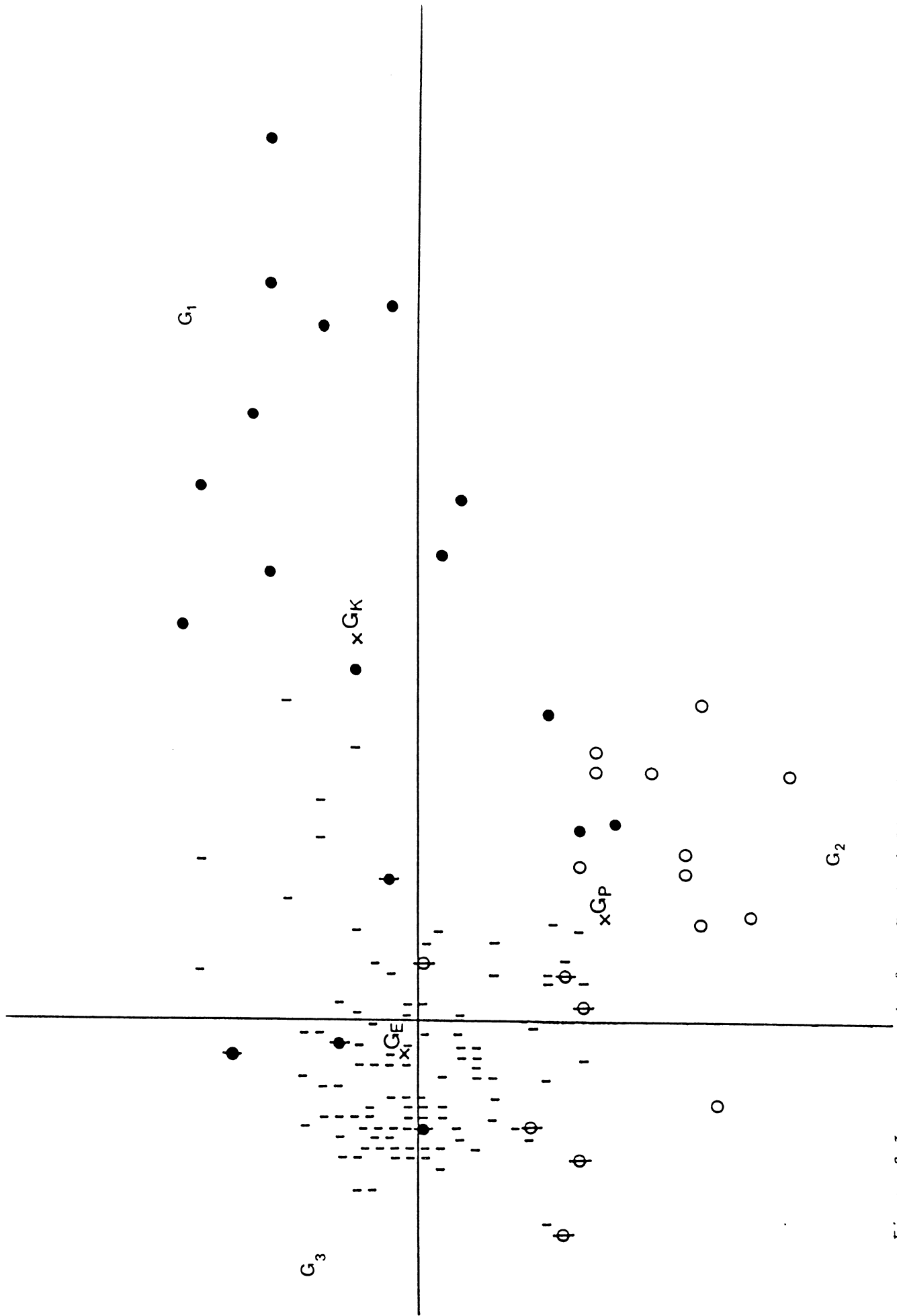


Figure 3.1

Analyse Factorielle Discriminante (41 variables, 169 patients)

| : Escherichia Coli o : Proteus • : Klebsielle
 ◐ : à porte d'entrée ◑ : à porte d'entrée

Ce graphique amène les observations suivantes :

- a. Les centres de gravité sont proches.
- b. Les groupes "Klebsielle" et "Proteus" sont très dispersés autour de leur centre de gravité.
- c. Le groupe "Escherichia Coli" est au contraire beaucoup plus homogène.
- d. Quatre "Klebsielles" et six "Proteus" sont projetés dans la zone de l'Escherichia Coli ou à proximité de celle-ci.
- e. Il existe un groupe de "Klebsielles" (G1) éloigné dans la direction de D1.

Le TABLEAU 3.IV donne les variables les plus fortement corrélées à chacune des deux composantes discriminantes.

VARIABLES	D1	VARIABLES	D2
B PNEUMOPATHIE	+ 0,55	INFECTION URINAIRE	+ 0,40
ACQUISITION REANIMATION	+ 0,56	ACQUISITION HOSPITALIERE	- 0,47
ACQUISITION HOSPITALIERE	+ 0,45		
PORTE ENTREE CATHETER	+ 0,43		
PORTE ENTREE URINAIRE	- 0,56		
INFECTION URINAIRE	- 0,48		

TABLEAU 3.IV

On trouve, dans le TABLEAU 3.V, le résultat du classement des patients obtenu en utilisant la distance de Mahalanobis.

GERMES CLASSEMENT	ECOLI	PROTEUS	KLEBSIELLE
ECOLI	116	4	4
PROTEUS	12	13	2
KLEBSIELLE	6	0	12

TABLEAU 3.V

Le pourcentage de "bien classés" est de 83,4%. Le graphique et les données de ces tableaux seront interprétés dans le paragraphe suivant.

3.1.3 Discussion et Conclusion

L'étude des corrélations entre les variables initiales et les composantes discriminantes D1 et D2 permet de mettre en évidence les variables qui interviennent le plus pour séparer les trois groupes. Le graphique doit donc s'interpréter en tenant compte des données du **TABLEAU 3.IV** :

- a. En projection sur l'axe D1, les groupes "Escherichia Coli" et "Proteus" ne se distinguent pas. D1 sépare un groupe G1, constitué de patients présentant une Klebsielle, de l'ensemble des autres patients. "B_peumopathie", "acquisition en réanimation" et "acquisition hospitalière" sont les variables les plus corrélées positivement à l'axe D1 (TABLEAU 3.IV). On met ainsi en évidence les caractéristiques moyennes des individus de G1 : patients présentant une septicémie due à une Klebsielle, avec des localisations septiques pulmonaires, et contractée à l'hôpital, en

particulier en réanimation. En examinant les données, on constate en effet que les 12 "Klebsielles" de G1 correspondent toutes à des septicémies nosocomiales ou à une localisation septique pulmonaire, 9 d'entre eux répondant aux deux critères.

- b. En projection sur la deuxième composante discriminante D2, on distingue un groupe G2, en majorité composé de "Proteus", tandis que les groupes "Ecoli" et "Klebsielle" se superposent. C'est essentiellement par l'absence d'infection urinaire et le lieu d'acquisition communautaire que ce groupe G2 diffère des autres patients. En particulier, aucun patient de G2 ne présente de septicémie à porte d'entrée urinaire.
- c. On observe, au centre du graphique, un groupe homogène G3. Ce groupe est en majorité composé de patients présentant une septicémie non contractée à l'hôpital, à porte d'entrée urinaire. On retrouve, dans ce groupe, la presque totalité des "Escherichia Coli", mais aussi quatre "Proteus" et quatre "Klebsielle". Ces huit patients souffrent d'une septicémie à porte d'entrée urinaire et ainsi, ne se distinguent pas de ceux qui présentent une septicémie causée par l'Escherichia Coli.

Le pourcentage de bien classés obtenu à partir de cette analyse (TABLEAU 3.V) doit être interprété en tenant compte de l'ensemble des résultats précédents. Notre objectif était de déterminer de nouvelles associations de signes, pertinentes pour différencier le Proteus et la Klebsielle de l'Escherichia Coli. Pour ce faire, nous avons étudié si les trois groupes "Escherichia Coli", "Klebsielle" et "Proteus", pouvaient être différenciés à partir de combinaisons linéaires des variables intervenant en prémisses des règles du groupe E-K-P.

L'analyse discriminante permet de construire les combinaisons linéaires de ces variables qui prennent les valeurs les plus différentes possibles pour ces trois groupes. Il apparaît que les variables discriminantes permettent plutôt de distinguer les trois groupes G1, G2 et G3. Les caractéristiques principales de ces groupes, septicémies nosocomiales à localisation(s) septique(s) pulmonaire(s) (G1), septicémies nosocomiales à porte d'entrée ni urinaire ni pulmonaire (G2), septicémie à porte d'entrée urinaire (G3), ne peuvent être considérées comme pertinentes pour l'identification du groupe E-K-P. En particulier, celles du groupe G1 sont proches des signes cliniques observés chez les patients qui souffrent d'une septicémie causée par le Pneumocoque ou le Staphylocoque. En créant un sous-but regroupant "Escherichia Coli", "Proteus", "Klebsielle", sous-but correctement atteint par le système expert, ces caractéristiques pourraient néanmoins être utilisées dans des règles de second niveau, destinées à séparer les trois groupes.

Cependant, notre étude a été réalisée sur un échantillon comportant un nombre peu important de septicémies causées par le Proteus ou la Klebsielle. Les résultats que nous avons obtenus devront donc être confirmés par une étude effectuée sur un échantillon plus large.

Ces différents résultats ont été fournis aux experts du groupe SES. Ils ne permettent pas, actuellement, de déterminer de nouvelles règles pertinentes pour le diagnostic différentiel entre les septicémies dues à l'Escherichia Coli, celles dues au Proteus et celles dues à la Klebsielle. Mais ils confirment l'intuition clinique des experts : les patients présentant un Escherichia Coli, un Proteus ou une Klebsielle à porte d'entrée urinaire ne sont pas distingués par

notre étude. Les signes cliniques des septicémies causées par le Proteus et l'Escherichia Coli nous apparaissent très proches, comme dans la base de connaissances (TABLEAU 3.I). Nous retrouvons la distinction entre les "Klebsielles" à localisation septique pulmonaire et les autres patients du groupe E-K-P (TABLEAU 3.II).

Nous avons ensuite poursuivi l'étude des possibilités de discrimination en analysant le cas des germes à faible prévalence.

3.2 Etude des germes à faible prévalence

Parmi les 25 germes que le système expert se propose d'identifier, certains sont très rarement rencontrés dans notre sélection (chapitre III A : TABLEAU 2.I). Ce résultat s'explique par la faible prévalence de ces bactéries dans les septicémies. On peut distinguer, parmi ces germes rares, deux groupes (TABLEAU 3.VI) :

GERMES	FREQUENCES (%)
GROUPE I	
ACINETOBACTER	1 (0,2%)
SERRATIA	3 (0,5%)
GROUPE II	
ANAEROBIES ¹	6 (1%)
BRUCELLE ET YERSINIA	3 (0,5%)
CANDIDA	4 (0,7%)
GONOCOQUE	1 (0,2%)

¹: BACTEROIDE et CLOSTRIDIUM

TABLEAU 3.VI

Le problème posé par le diagnostic des septicémies dues à la *Serratia* ou à l'*Acinetobacter* (groupe I) est du même type que celui étudié dans le paragraphe précédent : ces deux micro-organismes induisent des signes cliniques qui sont considérés par les experts comme proches de ceux observés chez les septicémies causées par la *Klebsielle* ou l'*Escherichia Coli*.

L'identification des six germes du deuxième groupe est de nature différente. Les infections causées par ces bactéries apparaissent en effet généralement dans des contextes très particuliers, dont nous donnerons deux exemples : chez les femmes après avortement septique, pour les septicémies à *Clostridium*, chez les patients présentant une altération profonde de l'état général, associée à une maladie sous-jacente grave et à la préexistence d'une agression médicamenteuse (antibiothérapie, corticothérapie), pour les septicémies à *Candida*. En dehors de situations particulières de ce type, les experts estiment que ces six germes sont très improbables et par conséquent ne les proposent pas en diagnostic de première intention.

Les septicémies causées par l'un de ces micro-organismes, observées dans notre sélection, sont difficilement reconnues par DIAGSES (chapitre III A : TABLEAU 2.III). Ce résultat, obtenu sur un très petit nombre de cas, ne peut pas être considéré comme représentatif de la capacité du système à identifier ces germes. Il doit cependant être expliqué et éventuellement corrigé. Notons que l'analyse des données ne peut pas actuellement apporter d'aide à l'identification de ces micro-organismes. en raison précisément de leur trop faible fréquence dans notre échantillon. Nous avons donc adopté, avec les experts, la démarche suivante :

- Etude des connaissances de DIAGSES relatives à ces six germes.
- Analyse des dossiers de la base de données non identifiés par DIAGSES.
- Possibilité de modifier la base de connaissances.

3.2.1 L'Expertise

Nous avons étudié les règles de DIAGSES qui permettent de déduire ces germes à faible prévalence. En reprenant la terminologie du paragraphe précédent, nous désignerons par signe une prémisse de l'une de ces règles, qui peut être une donnée clinique ou une association de données cliniques. Nous dirons qu'un signe S est spécifique (respectivement non spécifique) à un germe donné G, s'il permet, dans la base de connaissances, de déduire uniquement ce germe G (plusieurs germes G, G1, G2 ...) comme l'illustre la figure suivante :



Le TABLEAU 3.VII donne, pour chaque germe étudié, le nombre total de règles et le nombre d'entre elles qui utilisent des signes spécifiques :

GERMES	NOMBRE DE REGLES	SIGNES SPECIFIQUES
ANAEROBIES ¹	23	6
BRUCELLE ET YERSINIA	8	1
CANDIDA	9	2
GONOCOQUE	4	1

¹: BACTEROIDE et CLOSTRIDIUM

TABLEAU 3.VII

Les signes spécifiques

Pour chacun de ces signes spécifiques S_j , nous avons obtenu, sur les données de notre sélection, le tableau de fréquence $S_j \times D$ où D désigne la variable "diagnostic", c'est-à-dire le type de germe. Cette variable qualitative D peut donc prendre 25 valeurs différentes associées aux 25 germes possibles. On notera D_j la conclusion de la règle unique associée au signe spécifique S_j . Les résultats apparaissent dans le TABLEAU 3.VIII :

SIGNES SPECIFIQUES S_j	EFFECTIFS DES SIGNES S_j	CONCLUSIONS D_j DES REGLES ASSOCIEES (CV)	EFFECTIFS $S_j \times D_j$
1 ANEMIE HEMOLYTIQUE ET CHOC TI ET INSUFFISANCE RENALE	0	ANAEROBIES (0,9)	0
2 INFECTION DENT ET PLEURESIE PURULENTE	0	ANAEROBIES (0,8)	0
3 GANGRENE GAZEUSE	0	ANAEROBIES (1)	0
4 RHABDOMYOLYSE	0	ANAEROBIES (0,9)	0
5 PHLEGMON PERI AMYGDALIEN ET PLEURESIE PURULENTE	0	ANAEROBIES (0,8)	0
6 PHLEGMON PERI AMYGDALIEN ET ABCES DU POU MON	0	ANAEROBIES (0,8)	0
7 RISQUE PROFESSIONNEL = 2.00 (CONTACTS BOVINS-OVINS)	8	BRUCELLE (0,3)	0
8 TOXICOMANE ET CHORIORETINITE	0	CANDIDA (0,9)	0
9 TOXICOMANE ET UVEITE	0	CANDIDA (0,9)	0
10 ARTHRITE TEMPORO MAXILLAIRE	0	GONOCOQUE (0,8)	0

TABLEAU 3.VIII

Les TABLEAUX 3.VII et 3.VIII amènent les remarques suivantes :

- a. Pour chaque germe étudié, il existe des signes spécifiques. Les règles associées ont des coefficients de vraisemblance élevés. Ces règles expriment des heuristiques qui permettent d'identifier le germe causal dans des situations pathologiques très particulières : "Si le malade est en état de choc toxi-infectieux et présente une anémie hémolytique associée à une insuffisance rénale, alors la septicémie est très probablement due (.9) au Candida ".
- b. Ces signes spécifiques ne sont pas observés dans notre sélection. Les tableaux cliniques présentés par les malades du dossier BD_SES, qui souffrent d'un septicémie causée par l'un des germes étudiés, ne correspondent pas aux cas précis décrits par les heuristiques.

Les signes non spécifiques

Les autres signes qui permettent, dans la base de connaissances, de déduire ces germes rares, correspondent à des pathologies plus fréquemment rencontrées dans les septicémies (broncho-pneumopathie, endocardite...). Devant ces pathologies, les experts estiment que l'infection peut être due à l'un des six germes rares, mais qu'il faut envisager préférentiellement d'autres diagnostics. Chacun de ces signes intervient donc, dans la base de connaissances, en prémisse de plusieurs règles, les pondérations associées favorisant le diagnostic de germes plus probables, dans de tels contextes cliniques, que l'une des bactéries étudiées.

Par exemple, les trois règles :

SI broncho-pneumopathie ALORS Pneumocoque (.5),

SI broncho-pneumopathie ALORS Hemophilus (.5),

SI broncho-pneumopathie ALORS Anaérobie (.2),

expriment la connaissance experte "une septicémie associée à une broncho-pneumopathie est souvent due au Pneumocoque (CV= 0,5) ou à l'Hémophilus (CV= 0,5), éventuellement à l'Anaérobie (CV= 0,2)".

3.2.2 Les Cas observés dans notre Sélection

Rappelons que le système expert propose dans ses conclusions un, deux, le plus souvent trois germes possibles, par ordre décroissant de vraisemblance. LINXES fournit la liste des dossiers de la base de données pour lesquels le germe responsable de l'infection n'est pas identifié en première position par DIAGSES.

Les données cliniques des 14 patients de la base de données qui présentent une septicémie causée par l'Anaérobie, le Candida, le Gonocoque ou la Brucelle ont été analysées par les experts du GROUPE SES. L'objectif était double : déterminer si, au niveau actuel de l'expertise, le diagnostic était possible pour ces patients, et, en cas de réponse positive, déduire de l'examen de ces dossiers de nouvelles associations de signes, pertinentes pour l'identification de ces bactéries.

Les experts ont estimé qu'aucun de ces 14 patients ne présentait des signes cliniques permettant d'identifier, en première position, le germe réel. Pour deux dossiers, ils ont jugé que ce germe pouvait être proposé en troisième position. L'un de ces deux germes pathogènes (Anaérobie) avait été reconnu en troisième position par le

système expert. Les pathologies présentées par les autres patients n'ont pas été jugées significatives par les experts, pour évoquer le germe réel dans l'une des trois premières positions. Par ailleurs, les réponses du système pour ces 14 dossiers ont été semblables aux diagnostics alors proposés par les experts, comme en témoignent les deux exemples réels du TABLEAU 3.IX :

SIGNES CLINIQUES	DIAGNOSTICS
<p>PATIENT I</p> <p>conscience normale femme age =78 non contraction hospitaliere; glob.blancs_mille =13.5 pnn=91 plaquettes_mille=277 vs=78 temperature=39.9 insuffisance cardiaque lithiase biliaire non antibiotherapie anterieure porte entree colique</p>	<p>GERME REEL : ANAEROBIE.</p> <p>CONCLUSIONS DES EXPERTS: probablement l'escherichia coli ou la salmonelle, éventuellement l'anaérobie.</p> <p>DIAGSES :</p> <p>1) escherichia coli 2) salmonelle 3) anaérobie</p>
<p>PATIENT II</p> <p>conscience normale femme age =65 non contraction hospitaliere; glob.blancs_mille =8.4 pnn=50 plaquettes_mille=357 vs=55 temperature=37.7 broncho-pneumopathie splenectomie insuffisance cardiaque antibiotherapie anterieure</p>	<p>GERME REEL : CANDIDA.</p> <p>CONCLUSIONS DES EXPERTS: probablement le pneumocoque ou l'hémophilus.</p> <p>DIAGSES :</p> <p>1) pneumocoque 2) hemophilus 3) klebsielle</p>

TABLEAU 3.IX

Cette étude nous a permis de préciser la nature des difficultés posées par l'identification des septicémies causées par l'Anaérobie, le Candida, le Gonocoque ou la Brucelle et de vérifier la pertinence, au niveau actuel de l'expertise, des connaissances de DIAGSES relatives à ces bactéries.

Ces germes rares sont généralement associés à des situations pathologiques très particulières décrites par les heuristiques du TABLEAU 3.VIII, qui permettent alors de les identifier.

Ces signes cliniques spécifiques ne sont pas observés dans notre sélection, il n'y a donc pas de contradiction entre les heuristiques et les réponses du système.

En dehors des cas types décrits par les heuristiques, les experts estiment qu'il n'existe pas de signe clinique discriminant pour le diagnostic, en première position, de ces micro-organismes. Cette intuition clinique des experts a été vérifiée sur les septicémies du dossier BD_SES causées par l'une de ces bactéries.

450 dossiers supplémentaires sont archivés chaque année dans le dossier informatisé BS_SES. Cette base de cas de plus en plus importante sera utilisée pour améliorer les connaissances épidémiologiques relatives à ces bactéries et surveiller leur fréquence d'apparition dans les septicémies.

4 CONCLUSION

Dans cette étape, nous avons pu évaluer les performances globales, en termes statistiques, du Système Expert Septicémie. Nous

avons aussi mis en évidence des insuffisances de diagnostic, lorsque nous avons montré que le système échouait dans le diagnostic différentiel entre Escherichia Coli - Proteus - Klebsielle, ou qu'il avait des performances médiocres pour le diagnostic des germes rares.

Les analyses statistiques complémentaires ont démontré aussi que la discrimination entre germes n'était pas, dans tous les cas, chose facile, au vu des seuls signes cliniques ou biologiques.

Puisque le problème n'est pas décidable, on peut au moins se poser la question :

Est-il utile, pour le médecin, de savoir si le germe est Escherichia Coli ou Proteus ? A partir de quelles données cliniques et bactériologiques prescrit-il sa thérapeutique antibiotique ?

Nous verrons dans la prochain chapitre que les médecins ont modifié leur réflexion stratégique à partir des données que nous venons d'exposer.

CHAPITRE IV : - STRATEGIE D'AMELIORATION DES PERFORMANCES-

CHAPITRE IV : STRATEGIE D'AMELIORATION DES PERFORMANCES.

1 INTRODUCTION

Dans ce chapitre, nous exposons les différentes étapes qui ont conduit à améliorer de manière très significative les performances de **DIAGSES**. La stratégie d'amélioration a été développée sur les dossiers de la base de données **BD_SES** des années 1985-1986 (555 cas). Ce sont ceux que nous avons utilisés dans le précédent chapitre pour évaluer la base de connaissances originelle.

A chaque étape, la méthodologie suivante a été employée :

- a. Evaluation de la base de connaissances par **LINXES**.
- b. Analyse des résultats de cette évaluation avec les experts.
- c. Analyse des données : en fonction des résultats de l'évaluation, de la base de connaissances et en collaboration avec les experts.
- d. Modification de la base de connaissances : prise en compte des modifications proposées par les experts et /ou de celles suggérées par les connaissances numériques extraites de la base de données par les analyses statistiques.
- e. Retour à a) pour vérifier la pertinence des modifications et réitérer le processus.

Cette démarche est illustrée par la **FIGURE 1.I**.

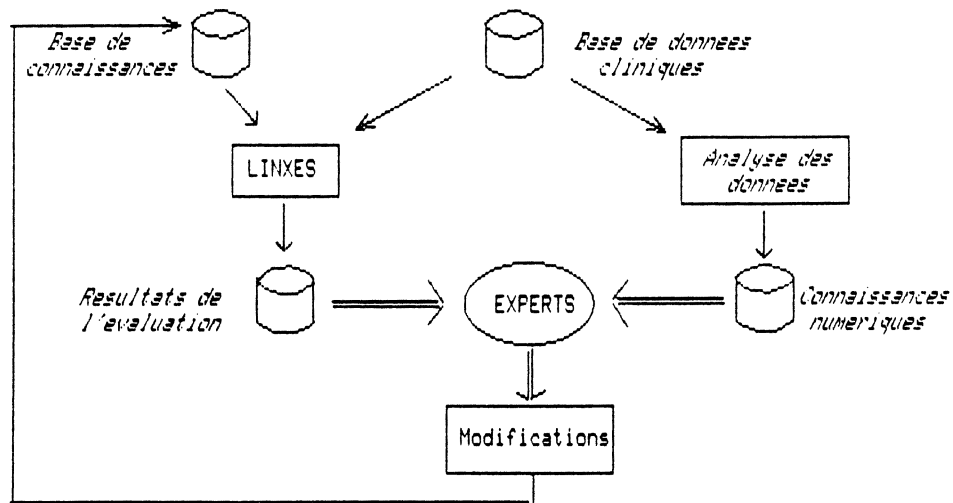


FIGURE 1.1

Le processus décrit par la figure précédente a été réitéré quatre fois.

Etape I : DIAGSES1.

Renouvellement de la connaissance : les Groupes Bactériens de Sensibilité Voisine (GBSV). Création de la nouvelle base de connaissances DIAGSES1, et évaluation sur les 555 dossiers (85-86).

Etape II : DIAGSES2.

Modification par les experts de DIAGSES1 sur la base des résultats de l'évaluation précédente. Evaluation de la nouvelle base de connaissances DIAGSES2.

Etape III : DIAGSES3.

Création d'une nouvelle base de connaissances DIAGSES3 prenant en compte les résultats d'analyses bivariées. Evaluation de cette base de connaissances.

Etape IV : DIAGSES4.

Prise en compte , dans le système expert, de scores construits à partir d'analyses factorielles discriminantes et évaluation du nouveau système DIAGSES4.

CONCLUSION : évaluation sur un ensemble de cas nouveaux (290 dossiers de l'année 1987) des cinq bases de connaissances étudiées (DIAGSES, DIAGSES1-4).

Dans la première étape, nous allons montrer comment l'ensemble des études réalisées précédemment nous ont amené à modifier la base de connaissances originelle de DIAGSES pour mieux répondre à l'objectif principal du système qui est l'antibiothérapie de première intention.

2 ETAPE I : LES GROUPES BACTERIENS DE SENSIBILITES VOISINES

2.1 Nécessité et définition

Le Système Expert Septicémie a été construit pour apporter une aide réelle au médecin confronté à une situation infectieuse grave. Nous savons que l'évolution favorable de la septicémie est liée à l'institution rapide d'un traitement antibiotique efficace. Ce traitement dépend du germe responsable de l'infection, de la localisation septique et de l'état général du patient. L'étape d'identification du germe est donc nécessaire.

Lorsque les résultats bactériologiques ne sont pas encore connus, le médecin doit déterminer à partir des seules données cliniques et des paramètres biologiques standards, le germe le plus probable et en déduire l'antibiothérapie de première intention adaptée.

L'une des finalités principales du Système Expert Septicémie est de l'aider à résoudre ce problème que nous savons complexe. En particulier, DIAGSES a été mis au point pour réaliser l'étape d'identification du germe causal, dans l'attente des résultats bactériologiques. L'utilisation du système dans ce cadre ainsi que les étapes réalisées sont illustrés par la FIGURE 2.I :

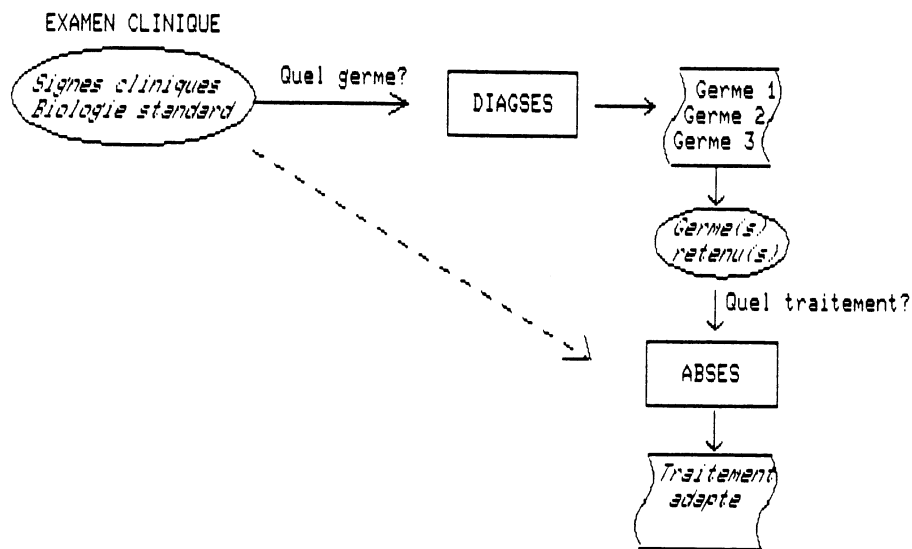


FIGURE 2.I

Cette démarche est équivalente à la démarche intellectuelle utilisée par les experts en maladies infectieuses confrontés à une telle situation d'urgence. Plus précisément, l'expertise doit considérer deux problèmes de complexité différente :

- a. Type I : les signes cliniques présentés par le malade apparaissent à l'expert suffisamment significatifs, compte tenu de ses connaissances et de son expérience, pour que le diagnostic s'impose avec une quasi-certitude. Il prescrira alors l'antibiotique à la fois efficace contre ce germe et adapté au patient.

b. Type II : la situation est plus douteuse. Il lui est difficile de se déterminer entre deux germes de "vraisemblances" équivalentes. Par exemple, une septicémie pour laquelle l'expert suspecte un Escherichia Coli ou un Proteus.

Nous allons voir comment ses connaissances en infectiologie et en antibiothérapie lui permettent, dans ces cas difficiles, d'instituer un traitement efficace et adapté.

L'existence de situation du type II a été confirmée par les résultats de l'évaluation précise de DIAGSES sur la base de données BD_SES et expliquent pour une part les erreurs de discrimination du système. Les études statistiques que nous avons effectuées n'ont pas apporté de solution à ces problèmes de diagnostic.

Dans certaines de ces situations difficiles, il est toutefois possible de prescrire un antibiotique efficace contre les germes suspectés et adapté au patient, tout en respectant l'écologie bactérienne du malade et de l'environnement ainsi que les contraintes économiques.

En l'absence de l'antibiotique idéal (à spectre très large et peu coûteux), nous savons que le choix d'un antibiotique est un problème complexe et repose sur la notion de spectre bactérien. Par exemple, tel antibiotique (PENI G) peut se révéler efficace pour un germe donné (PNEUMOCOQUE), mais sans effet ou même contre-indiqué pour un autre germe (ESCHERICHIA COLI). On trouve dans différentes monographies type "le point actuel sur les antibiotiques" [MOUTON 87], l'exposé de ces différentes notions.

Plusieurs bactéries ont une sensibilité voisine aux antibiotiques. Par exemple, l'Escherichia Coli et Le Proteus sont tous les deux sensibles aux mêmes antibiotiques, à un détail près. Ainsi, si l'ensemble des signes cliniques laisse suspecter un Escherichia Coli ou un Proteus et que la porte d'entrée du germe est urinaire, l'antibiotique adapté est généralement un "C3G" (Céphalosporine de Troisième Génération). Cette propriété ne concerne pas seulement certains des germes ayant des tableaux cliniques voisins : le Streptocoque B, qui n'est pas bien identifié par DIAGSES, peut être traité, en première intention, de la même manière que le Pneumocoque ou le Méningocoque, germes bien identifiés.

Dans sa conception initiale, DIAGSES se propose d'identifier le germe responsable de l'infection parmi les 25 micro-organismes retenus comme possibles par les experts en cas de septicémie. Ces 25 conclusions différentes du système correspondent à la classification bactériologique. Cette démarche est bien adaptée aux situations de type I. Les remarques précédentes montrent qu'il n'est pas nécessaire, au niveau de l'étape diagnostique réalisée par DIAGSES, de différencier l'ensemble de ces 25 germes, si l'on se place du point de vue de l'efficacité thérapeutique.

Ainsi, pour prendre en compte les situations de type II, c'est-à-dire quand le diagnostic différentiel entre plusieurs germes suspectés est très difficile à réaliser, et mieux adapter le système expert à sa vocation principale d'aide à la thérapeutique en situation d'urgence, les experts ont modifié les conclusions de DIAGSES pour déterminer :

Quels sont les groupes de germes différents du point de vue de

l'antibiothérapie de première intention ?

Les réponses à cette question ne sont plus 25 germes possibles mais dix classes regroupant les germes correspondant à une même antibiothérapie de première intention. Ces classes regroupent les germes ayant une sensibilité voisine aux antibiotiques, les Groupes Bactériens de Sensibilité Voisine (GBSV) [MOUTON 88].

La démarche ayant permis d'établir cette classification peut être représentée par la FIGURE 2.II. L'axe horizontal est orienté dans le sens de la résistance croissante aux antibiotiques. Partant de la classification bactériologique, cette figure montre comment les différents germes ont été affectés dans les différents GBSV.

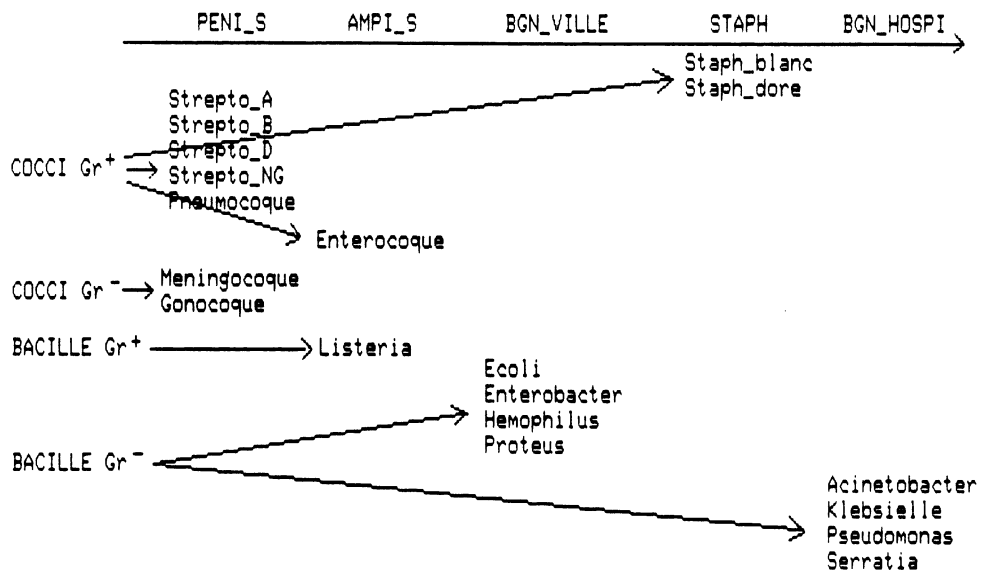


FIGURE 2.II

CHAPITRE IV:- STRATEGIE D'AMELIORATION DES PERFORMANCES -

Le TABLEAU 2.I donne la correspondance, établie par les experts, entre la classification GBSV et la classification bactériologique.

CLASSIFICATION BACTERIOLOGIQUE	GBSV
STAPHYLOCOQUE DORE STAPHYLOCOQUE BLANC	STAPHYLOCOQUES
GONOCOQUE MENINGOCOQUE PNEUMOCOQUE STREPTOCOQUE A STREPTOCOQUE B STREPTOCOQUE D STREPTOCOQUE NG	PENI_S
ENTEROCOQUE LISTERIA	AMPI_S
ESCHERICHIA COLI ENTEROBACTER HEMOPHILUS PROTEUS	BGN VILLE
SALMONELLE	BGN SALMONELLES
ACINETOBACTER KLEBSIELLE PSEUDOMONAS SERRATIA	BGN HOSPITALIERS
BRUCELLE YERSINIA	BGN BRUCELLE
BACTEROIDE CLOSTRIDIUM	ANAEROBIES
CANDIDA	CANDIDA
AUTRES GERMES	AUTRES GERMES

TABLEAU 2.I

ABSES utilise comme faits initiaux les données cliniques du patient et, naturellement, le germe responsable de l'infection. La base de connaissances de ABSES a donc été modifiée pour prendre en compte non plus les 25 germes possibles, mais les dix classes GBSV. Ce travail est actuellement terminé, mais dépassant le cadre de notre

étude, nous ne l'exposerons pas dans ce mémoire.

Par définition, les GBSV regroupent les germes qui ont une sensibilité voisine aux antibiotiques. La modification des buts de DIAGSES en classes GBSV et leur prise en compte dans la base de connaissances de ABSES, ne pénalise pas la qualité des conseils du système expert, lorsqu'il est utilisé dans ce cadre.

Cette étude nous a permis d'améliorer sensiblement les performances du système expert, comme nous le montrerons au paragraphe suivant. Elle met par ailleurs en évidence une double utilité :

- a. Utilité d'obtenir une évaluation précise des performances du système expert. Ce sont les résultats de cette évaluation qui ont permis de mettre en évidence les dangers de la démarche un patient - un germe - un antibiotique et donc la nécessité d'introduire les GBSV.
- b. Utilité de l'expertise : la notion de Groupe Bactérien de Sensibilité Voisine est directement liée aux connaissances expertes.

Avec les experts, nous avons adapté la base de connaissances de DIAGSES à l'identification des classes GBSV. Notre travail a ensuite consisté à évaluer cette nouvelle base de connaissances DIAGSES1 sur la base de données.

2.2 Construction de la base de connaissances GBSV

La base de connaissances (DIAGSES1) destinée à l'identification des Groupes Bactériens de Sensibilité Voisine est dérivée de celle de DIAGSES.

L'ensemble des signes cliniques et des paramètres biologiques retenus par les experts pour leur pertinence vis-à-vis du diagnostic des germes est demeuré inchangé. Chacun des 25 germes a été remplacé par sa classe GBSV d'appartenance.

Cette substitution a été réalisée de la manière suivante :

- a. Les règles de DIAGSES qui utilisent la même prémisse P pour déduire des germes "G1",..., "Gi",..., "Gk" appartenant à une même classe GBSV "GBSV_1", ont été regroupées en une seule règle (TABLEAU 2.II) :

R1: P EST_EN_FAVEUR G1 <CV1>

Ri: P EST_EN_FAVEUR Gi <CVi> --> R: P EST_EN_FAVEUR GBSV_1 <CV>

Rk: P EST_EN_FAVEUR Gk <CVk>

TABLEAU 2.II.

Le coefficient de vraisemblance CV a été évalué par les experts. Dans la plupart des cas, les experts ont choisi $CV = \text{Max}(CV_i)$, $i=1\dots k$, valeur qui correspondait à leur intuition clinique, selon la même démarche pragmatique adoptée lors de l'estimation des pondérations associées aux règles de DIAGSES (chapitre A).

- b. Il existe dans la base de connaissances de DIAGSES des règles négatives, c'est-à-dire des règles qui permettent d'exclure un germe donné selon un degré de confiance CV. Par exemple, la règle "PORTE_ENTREE_URINAIRE EST_EN_FAVEUR NON STREPTO_N_G <0.90>",

exprime la connaissance experte "si la porte d'entrée du germe est urinaire, la septicémie n'est très probablement pas causée (0,9) par un Streptocoque non groupable". Pour chacune de ces règles, deux cas sont à envisager :

b.1 La prémisse correspondante permet par ailleurs de déduire positivement un (ou plusieurs) germe(s) appartenant à la même classe GBSV. Par exemple, une endocardite est en défaveur (.8) d'une septicémie à Méningocoque, mais en faveur d'une septicémie à Streptocoque D (.3) ou Streptocoque non groupable (.6), ces trois germes appartenant au groupe des PENI_S :

R1 : ENDOCARDITE EST_EN_FAVEUR STREPTO_N_G <.6>

R2 : ENDOCARDITE EST_EN_FAVEUR STREPTO_D <.3>

R3 : ENDOCARDITE EST_EN_FAVEUR NON MENINGOCOQUE <.8>.

Dans de tels cas, la règle à conclusion négative (R3) a été supprimée.

b.2 Dans le cas contraire, la règle négative a été conservée :

PNEUMONIE_ALVEOLAIRE EST_EN_FAVEUR NON SALMONELLE <.9>

ENDOCARDITE EST_EN_FAVEUR NON ANAEROBIES <.9>

La base de connaissances ainsi construite a été évaluée sur la base de données BD_SES en utilisant LINXES.

2.3 Evaluation de DIAGSES1

Les objectifs et la méthodologie de l'évaluation de la base de connaissances de DIAGSES1 sont identiques à ceux employés lors de l'évaluation de base de connaissances originelle. Les résultats selon les années 1985/86 étant comparables, nous présentons ceux obtenus à partir de la totalité des 555 dossiers de notre sélection.

Rappelons que la note 1 (respectivement 2) est attribuée lorsque le GBSV est identifié en première position (respectivement dans l'une des trois premières positions).

Nous avons reproduit dans le TABLEAU 2.III les fréquences des classes GBSV ainsi que leurs notes respectives.

GBSV	EFFECTIFS (%)	EFFECTIFS NOTE 1 (%)	EFFECTIFS NOTE 2 (%)
STAPH	125 (22,5%)	75 (60%)	115 (92%)
PENI S	153 (27,6%)	99 (64,7%)	147 (96%)
AMPI S	36 (6,5%)	0	8 (22%)
BGN VILLE	164 (29,6%)	126 (76,8%)	160 (98%)
BGN SALMONELLE	26 (4,7%)	2	12 (46%)
BGN HOSPITALIERS	34 (6,1%)	5 (15%)	18 (53%)
AUTRES GROUPES	17 (3%)	0	2

TABLEAU 2.III

Le pourcentage global d'identification des GBSV en première position (pourcentage de bien classés) est 55,3%, et il est de 83% dans l'une des trois premières positions.

L'analyse des résultats montre que les réponses du système sont comparables, du point de vue qualitatif, à celles de DIAGSES :

CHAPITRE IV:- STRATEGIE D'AMELIORATION DES PERFORMANCES -

- a. Les groupes les plus fréquents ("STAPH", "PENI_S", "BGN_VILLE") sont bien reconnus par le système dès la première position. Les groupes "AMPI_S", "SALMONELLE" et "BGN_HOSPITALIERS" sont plus difficilement reconnus. Les résultats concernant les GBSV regroupant les germes à faible prévalence sont semblables à ceux de DIAGSES.
- b. Le tableau de fréquences (GBSV_REELS) \times (PREMIERE CONCLUSION) permet d'étudier la qualité de la discrimination en première position des groupes par le système. Nous avons reproduit, dans le TABLEAU 2.IV, les résultats concernant les GBSV les plus fréquents.

DIAGSES1 GBSV REELS	EFFECTIFS						
	STAPH	PENI_S	AMPI_S	BGN_VILLE	BGN SALMO.	BGN HOSPI.	AUTRES
STAPH	75	24	0	19	0	2	5
PENI_S	26	99	0	26	0	0	2
AMPI_S	9	19	0	8	0	0	0
BGN_VILLE	25	8	0	126	0	0	5
SALMONELLE	1	2	0	20	2	0	1
BGN_HOSPI.	4	3	0	21	0	5	1
AUTRES	5	2	0	10	0	0	0

TABLEAU 2.IV

Les résultats de ce tableau amènent les remarques suivantes :

- a. Les "PENI_S" (respectivement "STAPH") sont bien reconnus par DIAGSES1. Lorsqu'ils ne sont pas identifiés en première position, ces germes sont classés dans les mêmes proportions dans les groupes "STAPH" et "BGN_VILLE" (respectivement "PENI_S" et "BGN_VILLE").
- b. Les "AMPI_S" sont des Streptocoques Entérocoques. Ces germes sont souvent classés en première position, de manière erronée, dans le groupe des "PENI_S" (Streptocoques "A", "B", "D" ou "Non Groupables").
- c. L'Escherichia Coli appartient au groupe "BGN_VILLE". On retrouve donc les confusions "SALMONELLE" - "BGN_VILLE" et "BGN_HOSPITALIER" - "BGN_VILLE" relevées lors de l'évaluation de DIAGSES.
- d. Le groupe "BGN_VILLE" est particulièrement bien reconnu par le système dès la première position (77%). La fréquence de ce groupe est cependant surestimée par DIAGSES1 dans des proportions comparables à celle de l'Escherichia Coli par DIAGSES.

Les performances du système DIAGSES1 pour l'identification des groupes GBSV en première position (respectivement dans l'une des trois premières positions) sont donc supérieures de 7% (respectivement de 17%) à celles de DIAGSES. Cette augmentation était attendue. Elle a pour conséquence une réelle amélioration de la qualité des conseils thérapeutiques fournis par le système expert.

La base de connaissances DIAGSES1 a été construite à partir des connaissances manipulées par DIAGSES. Les difficultés d'identification et les erreurs de discrimination qui subsistent sont donc comparables à celles mises en évidence lors de l'évaluation de DIAGSES.

L'ensemble de ces résultats a été analysé par les cliniciens du **GROUPE SES**. Les experts ont alors modifié empiriquement la base de connaissances de DIAGSES1 dans le but de corriger les erreurs mises en évidence par cette étude. Une nouvelle base de connaissances, DIAGSES2 a alors été mise au point.

3 **ETAPE II : CONSTRUCTION ET EVALUATION DE DIAGSES2**

3.1 Construction de la base DIAGSES2

Les difficultés discutées au paragraphe précédent ont amené les experts à réviser les règles de la base de connaissances de DIAGSES1. Trois cliniciens du groupe **SES** ont été chargés d'étudier les règles de la base de connaissances. Chacun d'entre eux a revu les règles qui correspondaient à ses compétences particulières : septicémies à Staphylocoque, septicémies à Escherichia Coli ("BGN_VILLE")... Une mise en commun des modifications proposées a ensuite été opérée, de façon à obtenir un consensus. Cette démarche est équivalente à celle qui avait été employée pour la mise au point du système originel DIAGSES.

Les objectifs des experts étaient, en particulier :

- a. augmenter la spécificité du système pour le diagnostic du groupe "BGN_VILLE", de façon à obtenir une meilleure identification des groupes "BGN_HOSPITALIER" et "SALMONELLE".
- b. améliorer le pourcentage d'identification du groupe "AMPI_S",

3.1.1 Corrections liées à la surestimation du groupe "BGN_VILLE"

La fréquence du groupe "BGN_VILLE" est surestimée, en première position, d'environ 10% par le système expert. Cette surestimation explique la plupart des difficultés rencontrées pour le diagnostic des GBSV en première position. En particulier, parmi les 26 Salmonelles, 20 (75%) sont classées à tort dans le groupe BGN_VILLE et cette proportion est de 60% pour les germes du groupe "BGN_HOSPITALIER". Ces résultats ont amené les experts à remettre en doute le bien fondé d'un certain nombre de règles qui concluent, dans DIAGSES1, au groupe "BGN_VILLE" ou à en modifier les pondérations.

Rappelons que la surestimation, dans les réponses du système, de la fréquence du groupe "BGN_VILLE" est consécutive à celle de l'Escherichia Coli (82% de l'effectif du groupe "BGN_VILLE") par la base originelle DIAGSES.

La méthodologie rigoureuse employée pour construire cette base de connaissances a été décrite au chapitre III A. En particulier, les règles ont été élaborées de façon à prendre en compte deux points de vue :

- Etant donné un germe pathogène, quels sont les signes cliniques les plus souvent rencontrés chez les patients qui souffrent d'une septicémie causée par ce germe ?

- Etant donné un signe ou une association de signes cliniques, quels sont les germes qui peuvent être liés à l'occurrence de ces signes ?

Cette prise en compte de l'expression duale de la connaissance a permis de construire une base de connaissances précise et complète.

L'Escherichia-Coli est le germe le plus souvent isolé dans les septicémies. Sa prévalence est de 23% dans notre sélection, fréquence comparable à celle retrouvée par de nombreux auteurs (chapitre III A).

Il semble qu'une prise en compte trop importante de cette prévalence élevée dans l'écriture des règles et dans l'évaluation des coefficients de vraisemblance soit à l'origine du problème étudié.

La fréquence de l'Escherichia Coli est non seulement prépondérante dans les septicémies mais, de plus, les experts estiment qu'un grand nombre de signes cliniques liés à la septicémie et utilisés dans la base de connaissances pour déduire le diagnostic, sont rencontrés, plus ou moins fréquemment, chez les patients qui souffrent d'une infection causée par l'Escherichia Coli. En d'autres termes, la réponse à la question "face à ce signe clinique, quels sont les germes **possibles** ?" (et non pas les plus probables), contient très souvent l'Escherichia Coli. Cette intuition clinique des experts sera confirmée à l'étape III, lorsque nous étudierons, pour chaque signe ou association de signes cliniques intervenant en prémisses des règles, les fréquences d'apparition des différents germes à partir des observations du dossier informatisé.

Cette donnée épidémiologique avait conduit les experts à introduire des règles qu'ils ont jugées, a posteriori, non réellement nécessaires pour le diagnostic de l'Escherichia Coli et qui n'étaient pas non plus des règles de conduite à tenir (exemple 1). Cette fréquence prépondérante de l'Escherichia Coli les avait aussi incités à affecter aux règles qui déduisent cette bactérie des pondérations qu'ils ont jugées, au cours de cette étude, trop élevées (exemple 2).

Exemple 1 :

Les experts estiment que l'association "arthrite" et "prothèse osseuse" est très en faveur d'une septicémie causée par le Staphylocoque, mais ils jugent qu'il est possible de retrouver une telle association chez un patient infecté par l'Escherichia Coli. Ce jugement s'était traduit par les règles

(R1) ARTHRITE ET PROTHESE_OS EST_EN_FAVEUR STAPH_AUREUS <0.90> ,

(R2) ARTHRITE ET PROTHESE_OS EST_EN_FAVEUR E_COLI <0.30> .

Il leur est apparu, au cours de cette étude, que, s'il était effectivement possible de retrouver une telle association chez une septicémie à Escherichia Coli, ces signes cliniques ne devaient cependant pas être pris en compte au niveau de la décision diagnostique, pour déduire l'Escherichia Coli.

Exemple 2 :

Considérons le patient réel suivant, qui souffrait d'une septicémie à Staphylocoque, et pour lequel le diagnostic du système expert (fourni par LINXES) avait été "BGN_VILLE" :

CONSCIENCE NORMALE;
FEMME;
AGE_ANS=54.00;
ACQUISITION CHIRURGIE;
CREATININE=80;
G.B MILLE=7.2;
PNN=79;
TEMPERATURE=39.2 ;
THROMBOPHLEBITE_SUPPUREE;
TUMEUR SOLIDE;
NON ANTIBIOTHERAPIE ANTERIEURE;
PORTE_ENTREE_CATHETER_IV;

Les experts ont diagnostiqué en première intention le Staphylocoque principalement en raison des signes "Thrombophlébite suppurée" et "Porte d'entrée Catheter", d'ailleurs utilisés dans des règles pour déduire "STAPH". Les pondérations des règles utilisant en prémisse "Acquisition chirurgie" et "Thrombophlébite suppurée" pour déduire "BGN_VILLE" ont alors été jugées, a posteriori, trop élevées. Ceci explique l'erreur du système.

L'étude de la base de connaissances originelle DIAGSES montre que les règles qui concluent à l'Escherichia Coli sont les plus nombreuses (47) et que ces règles sont souvent affectées de pondérations élevées. Nous retrouvons naturellement ces règles dans la base de connaissances de DIAGSES1 pour le diagnostic du groupe "BGN_VILLE".

Sur la base des résultats de notre évaluation et en étudiant les dossiers "mal classés" par le système expert fournis par LINXES, les experts ont réexaminé chaque signe (ou association de signes) cliniques intervenant en prémisse de ces règles, de façon à répondre aux questions suivantes :

- 1) "tel signe clinique est-il réellement en faveur de l'Escherichia Coli ou du groupe BGN_VILLE ?" sinon,
- 2) "est-il nécessaire de citer l'Escherichia Coli en présence de ce signe ?"
- 3) si oui à 1) ou 2) "avec quelle vraisemblance, par comparaison avec les autres hypothèses diagnostiques ?".

L'objectif était d'éliminer les règles jugées, de ce point de vue, non pertinentes pour le diagnostic du groupe "BGN_VILLE", et de réajuster les coefficients de vraisemblance pour obtenir une meilleure identification des GBSV moins fréquents que ce groupe dans les septicémies.

Cette démarche a permis d'apporter à la base de connaissances de DIAGSES1 un nombre important de modifications. 60 règles permettaient, dans DIAGSES1, de déduire "BGN_VILLE". Dix règles ont été supprimées (en particulier la règle R2 de l'exemple précédent). Dix coefficients de vraisemblance ont été diminués, en particulier ceux des règles faisant intervenir en prémisse le lieu d'acquisition de l'infection (chirurgie, médecine), comme le montre l'exemple suivant :

(R3) "ACQUISITION EN CHIRURGIE EST EN FAVEUR BGN_VILLE <.2>" (au lieu de <.4>).

Quatre règles négatives (en défaveur du groupe "BGN_VILLE") ont été introduites, comme les règles :

(R4) "ACQUISITION REANIMATION EST EN FAVEUR NON BGN_VILLE <.5>" et
"GRANULOPENIE EST EN FAVEUR NON BGN_VILLE <.5>",
signes qui permettent par ailleurs de déduire "BGN_HOSPITALIER".

3.1.2 Autres corrections consécutives aux résultats de l'évaluation

Les experts ont jugé nécessaire d'introduire de nouvelles règles permettant de décrire plus précisément les septicémies causées par le groupe "AMPI_S". Par exemple,

"INFECTION URINAIRE EST EN FAVEUR AMPI_S <.5>",

"SONDE URINAIRE EST EN FAVEUR AMPI_S <.5>".

Parallèlement, des corrections ont été apportées pour améliorer la discrimination entre les groupes "STAPH" et "PENI_S". Par exemple, les experts estiment qu'une septicémie déclarée chez un patient ayant bénéficié, antérieurement à cette maladie, d'une antibiothérapie, est plutôt en défaveur d'une infection due à un Streptocoque non entérocoque. Ce jugement n'était pas explicité dans la base de connaissances originelle. La règle :

"ANTIBIOTHERAPIE ANTERIEURE EST FAVEUR NON PENI_S <.3>"

a ainsi été introduite.

Les règles qui permettent de déduire le groupe "BGN_HOSPITALIER" ont été soigneusement étudiées. Des pondérations ont été modifiées, comme celles des règles suivantes :

(R5) "ACQUISITION REANIMATION EST EN FAVEUR BGN_HOSPITALIER <.7> " (au lieu de <.4>).

(R6) "INFECTION URINAIRE EST EN FAVEUR BGN_HOSPITALIER <.3> " (au lieu de <.6>). D'autres règles ont été ajoutées, comme la suivante :

"LESION COLORECTALE EST EN FAVEUR BGN_HOSPITALIER <.1>"

On remarquera que certaines des modifications précédentes vont dans le sens des résultats obtenus à partir de l'analyse multifactorielle précédemment réalisée sur les groupes "Escherichia Coli", "Proteus" et "Klebsielle". Les septicémies à infection urinaire étaient, en très grande majorité, causées par l'Escherichia Coli (R6). Le signe "ACQUISITION REANIMATION" y était prépondérant pour séparer les Klebsielles des autres groupes (R4) et (R5). Le lieu d'acquisition communautaire était un caractère important pour la distinction du groupe "Escherichia Coli" (R3) et (R4).

Parallèlement à ces corrections liées aux résultats de l'évaluation, de nouvelles règles ont été construites pour prendre en compte l'évolution des pathologies liées à la septicémie. En particulier, des règles concernant le Sida ont été introduites, par exemple :

"SIDA EST EN FAVEUR BGN_VILLE <.4>"

"SIDA EST EN FAVEUR MYCOBACTERIE <.2>" (Groupe "AUTRES GERMES").

L'ensemble des modifications réalisées figure dans les tableaux de l'annexe II qui montrent les corrections successives apportées à la base de connaissances de DIAGSES1 au cours de notre étude.

La nouvelle base de connaissances intégrant ces modifications issues de l'expertise, DIAGSES2, a été évaluée par LINXES sur la base de données BD_SES.

3.2 Evaluation de DIAGSES2

Le TABLEAU 3.I donne les pourcentages d'identification en première position des différentes classes GBSV par le système expert DIAGSES2. Ces résultats concernent les deux années 85 et 86 (555 dossiers).

GBSV	EFFECTIFS (%)	EFFECTIFS NOTE 1 (%)	EFFECTIFS NOTE 2 (%)
STAPH	125 (22,5%)	79 (63%)	110 (88%)
PENI_S	153 (27,6%)	112 (73%)	145 (95%)
AMPI_S	36 (6,5%)	1	19 (53%)
BGN_VILLE	164 (29,6%)	104 (64%)	145 (88%)
BGN_SALMONELLE	26 (4,7%)	4	13 (50%)
BGN_HOSPITALIERS	34 (6,1%)	10 (30%)	23 (68%)
AUTRES GROUPES	17 (3%)	0	5 (30%)

TABLEAU 3.I

55,85% des classes GBSV sont reconnues dès la première position par le système, alors que 83% de ces groupes bactériens sont correctement identifiés dans l'une des trois premières positions.

Ces pourcentages globaux sont très proches de ceux obtenus par la base de connaissances précédente (DIAGSES1). L'examen des performances pour chaque GBSV montre cependant qu'il existe des différences importantes entre les réponses des deux systèmes experts :

- a. Le groupe des "PENI_S", déjà bien identifié par DIAGSES1, est encore mieux reconnu en première position (+8%).
- b. Les performances du système concernant le groupe "BGN_HOSPITALIER" sont sensiblement améliorées.

- c. Le groupe "AMPI_S" est maintenant assez bien reconnu dans l'une des trois premières positions.
- d. Le pourcentage de bonnes réponses en première position pour le groupe "BGN_VILLE" diminue d'environ 10%.

Le tableau de fréquences (GBSV_REELS)x(PREMIERE CONCLUSION) reproduit ci-dessous permet d'évaluer la qualité des réponses de DIAGSES2.

DIAGSES2 (Conclusion N*1) GBSV REELS	EFFECTIFS						
	STAPH	PENI_S	AMPI_S	BGN_VILLE	BGN SALMO.	BGN HOSPI.	AUTR
STAPH	79	25	1	2	0	13	
PENI_S	25	112	1	4	0	7	
AMPI_S	9	19	1	4	0	7	
BGN_VILLE	25	19	1	104	2	8	
SALMONELLE	1	13	1	4	4	1	
BGN_HOSPI.	8	6	0	7	0	10	
AUTRES	6	7	1	2	0	0	

TABLEAU 3.II

On peut alors faire les remarques principales suivantes :

- a. la discrimination réalisée entre les groupes "STAPH" et "PENI_S" est semblable à celle obtenue avec la précédente base de connaissances DIAGSES1. Les erreurs d'identification concernant le groupe "STAPH" se partagent maintenant entre les groupes "PENI_S" et "BGN_HOSPITALIER", au lieu de "PENI_S" et "BGN_VILLE" dans DIAGSES1. C'est cohérent, puisque l'on a diminué le "poids" du groupe "BGN_VILLE" dans la base de connaissances et que l'on

sait que les signes cliniques qui permettent de déduire "BGN_HOSPITALIER" sont proches de ceux qui permettent de déduire "BGN_VILLE".

- b. les "PENI_S" sont très bien classés dès la première position. Seules subsistent les erreurs de discrimination avec le groupe "STAPH" déjà relevées pour DIAGSES1. On note que les "PENI_S" qui étaient affectés de manière erronée dans le groupe "BGN_VILLE" (26) sont, pour moitié d'entre eux, correctement reclassés.
- c. mais, parallèlement, les "BGN_VILLE" "perdus" sont classés à tort dans le groupe "PENI_S".

Ces résultats apparaissent, au niveau des performances globales, décevants. Il semble que les corrections opérées sur les règles de DIAGSES1 qui permettaient de déduire "BGN_VILLE" aient été trop importantes, en termes de suppression de règles et de diminution des valeurs de coefficients de vraisemblance, puisque la fréquence de ce groupe est maintenant sous-estimée.

Du point de vue thérapeutique, les experts ont néanmoins estimé que la qualité des réponses de cette base de connaissances DIAGSES2 était meilleure que celle de la précédente DIAGSES1. A taux de réponses correctes égaux, ce sont maintenant les germes les plus difficiles à traiter qui sont mieux reconnus par le système. Par exemple, la meilleure identification des "BGN_HOSPITALIERS", germes ayant une résistance importante aux antibiotiques, a été jugée positive.

4 **ETAPE III : DIAGSES3 : PRISE EN COMPTE DES RESULTATS DES ANALYSES BIVARIEES**

L'étude précédente marque les limites de la méthodologie qui a été employée pour améliorer les performances du système.

Cette méthodologie, basée sur une évaluation précise de l'aide clinique apportée par le système expert et ne prenant en compte que les connaissances expertes, avait déjà été utilisée avec succès pour mettre au point la base de connaissances originelle (chapitre III A).

Il semble que l'on soit parvenu à un point de saturation, à partir duquel toute modification entraînant une amélioration pour une ou plusieurs classes GBSV se fait au détriment de l'identification d'autres groupes.

Il nous est donc apparu nécessaire de ré-étudier la cohérence de la base de connaissances dans son ensemble et de confronter les règles de la base de connaissances aux observations de septicémie du dossier informatisé BD_SES.

4.1 **Création de DIAGSES3**

4.1.1 **Méthode**

Nous avons appliqué à la base de connaissances précédemment étudiée et évaluée, la méthodologie décrite au chapitre I.5.5. Nous allons voir que la confrontation des règles de la base de connaissances avec les observations de la base de cas nous a amenés à proposer de nombreuses modifications qui ont permis, après avoir reçu l'aval des experts, d'augmenter sensiblement les performances diagnostiques de notre système expert.

Dans la base de connaissances destinée à l'identification des GBSV, le raisonnement incertain n'affecte qu'une seule étape du raisonnement. Une première étape permet de structurer les signes de plus bas niveau en syndromes. Ce sont ensuite des associations de syndromes (et/ou de signes) qui permettent d'évoquer les probabilités bactériennes.

Désignons par $P = (P_k, k = 1, n)$, l'ensemble des prémisses des règles incertaines de la base de connaissances. P_k est donc un signe clinique ou un syndrome, ou une association de syndromes (et/ou de signes). Soit D la variable qualitative diagnostic, dont les modalités sont les dix classes GBSV possibles.

Nous avons étudié, par des analyses bivariées, le comportement individuel de chaque prémisse P_k des règles incertaines de la base de connaissances vis-à-vis du diagnostic D . Ces prémisses P_k ont été croisées avec la variable D sur les données de notre échantillon d'apprentissage (555 dossiers des années 85-86).

4.1.1.1 Données statistiques obtenues sur la base de cas

Il s'agissait tout d'abord d'obtenir les statistiques suivantes :

- nombre de cas pour lesquels on observe P_k ,
- nombre de cas (fk_j) pour lesquels on observe l'association (P_k et D_j),
- calcul des valeurs prédictives de P_k , pour chaque diagnostic D_j , observées sur la base de cas : $VPP_{kj} = P(D_j/E_k)$ ($j = 1, 10$),
- calcul des coefficients de vraisemblance "observés", CV_{P_kj} ($j = 1, 10$) à partir de la formule de SHORTLIFFE :

$$(CV_Pkj) \left\{ \begin{array}{l} CVO = 1 \text{ si } P(Dj) = 1, \\ \quad = [P(Dj/Ek) - P(Dj)] / [1 - P(Dj)] \text{ si } P(Dj/Ek) > P(Dj), \\ \quad = 0 \text{ sinon.} \\ \\ CVN = 1 \text{ si } P(Dj) = 0, \\ \quad = [P(Dj) - P(Dj/Ek)] / P(Dj) \text{ si } P(Dj/Ek) < P(Dj), \\ \quad = 0 \text{ sinon.} \end{array} \right.$$

4.1.1.2 Prise en compte de signes supplémentaires

Dans la base de données, il existe des items qui ne sont pas pris en compte par le système expert. Avec les experts, nous avons intégré à notre étude certains de ces signes ou syndromes qui n'étaient utilisés en prémisse d'aucune règle : "corticoïde", "pathologie colorectale"...

4.1.1.3 Construction du Tableau T = (P x D)

Nous avons alors construit le tableau complexe T (FIGURE 4.I), T = (P x D), P désignant l'ensemble des prémisses des règles incertaines et D la variable qualitative "diagnostic".

DIAGNOSTICS		D1	Dj	Dp
		p1	pj	pp
PREMISSSES				
P1	f1	f11 VPP11 CV_P11	f1j VPP1j CV_P1j	f1p VPP1p CV_P1p
.	.			
Pk	fk	fk1 VPPk1 CV_Pk1	fkj VPPkj CV_Pkj	fkp VPPkp CV_Pkp
.	.			
Pn	fn	fn1 VPPn1 CV_Pn1	fnj VPPnj CV_Pnj	fnp VPPnp CV_Pnp

fk = fréquence de la prémisse Pk sur la base de cas ;
 fkj =nombre de cas pour lesquels on observe Pk et Dj ;
 VPPkj = $P(Dj/Ek)$; pj = $P(Dj)$.

FIGURE 4.I, TABLEAU T : analyses bivariées des prémisses.

4.1.1.4 Utilisation du tableau T

Nous avons utilisé les données de ce tableau pour proposer des modifications de la base de connaissances de la manière suivante :

- a. **Les règles jamais déclenchées** : elles utilisent des prémisses qui ne sont pas observées dans notre population. Ces règles peuvent décrire des cas non observables et donc être inutiles. Elles peuvent aussi être trop "spécialisées", c'est-à-dire que les conditions requises pour les déclencher peuvent avoir été surévaluées par les experts. L'ensemble de ces règles a été déterminé et fourni aux experts.

- b. Nous avons considéré toutes les règles "positives" (en faveur du diagnostic) de la base de connaissances de la forme

(Rki) : $P_{ki} \longrightarrow D_i \langle CV_{0ki} \rangle$.

Pour ces règles, nous avons comparé le coefficient de vraisemblance CV_{0ki} qui avait été déterminé par les experts à celui calculé à partir de la formule de SHORTLIFFE CV_{Pki} . En cas de discordance, deux cas pouvaient se produire :

- si $VPP_{ki} < p_i$, la règle était en contradiction avec les observations de la base de cas. Nous avons proposé sa suppression.

- si $VPP_{ki} > p_i$, nous avons proposé de rétenir CV_{Pki} comme nouveau coefficient de vraisemblance de la règle (Rki).

- c. Une démarche équivalente a été réalisée pour les règles négatives existant dans la base de connaissances.

- d. Propositions de règles positives : Soit P_l une prémisse du tableau T et D_m un diagnostic, pour lequel il n'existe pas de règle en faveur de D_m

(R_{lm}) $P_l \longrightarrow D_m$.

Si $CV_{P_{lm}} > 0,1$, nous avons proposé d'introduire la règle (R_{lm}) dans la base de connaissances.

- e. Propositions de règles permettant d'exclure des diagnostics : En reprenant les mêmes notations, si $f_{lm} = 0$ (l'association du symptôme P_l et du diagnostic D_m n'est retrouvée chez aucun patient de notre sélection), nous avons proposé la règle "négative" :

" $P_l \longrightarrow \text{NON } D_m \langle .9 \rangle$ ".

4.1.1.5 Prise en compte des règles dépendantes

Nous avons été amenés à prendre en compte une limitation du moteur d'inférence SAM (qui existe aussi dans MYCIN). Considérons les deux règles :

(R1) (E1 et E2) \longrightarrow D1 <p> ,

(R2) E1 \longrightarrow D1 <q> .

Cette situation n'est pas rare. Par exemple, si $p > q$, ces règles peuvent s'interpréter comme suit : la présence du signe E1 permet d'évoquer D1 ; en présence de l'association E1 et E2, D1 devient très probable, tandis que E2 n'est pas pertinent pour le diagnostic D1.

Cependant, si les signes E1 et E2 sont présents dans la base de faits, SAM déclenchera les deux règles, affectant alors $p+q-pq$ à la vraisemblance de D1 (chapitre III A). En effet, SAM ne possède pas de "meta-règles" [VOYER 87] permettant de gérer de telles redondances.

Dans ces conditions, les experts, conscients de ce problème, avaient corrigé les effets de telles redondances en diminuant artificiellement le CV des règles du type R1 : dans notre exemple, le coefficient de vraisemblance corrigé de la règle R1 serait $(p-q)/(1-q)$, pour tenir compte de la dépendance des règles R1 et R2.

L'ensemble de ces règles clairement dépendantes a été déterminé et étudié séparément, de façon à

- a. vérifier leur pertinence : dans notre exemple, l'association "E1 et E2" apporte-t-elle réellement une information différente de celle apportée par le fait E1 seul ? En d'autres termes, a-t-on $P[D/(E1 \text{ et } E2)] > P[D/E1]$?

- b. En cas de réponse positive, opérer la correction précédente sur le CV_P calculé à partir des fréquences.

4.1.2 Résultats

A partir des données du tableau T, un nombre important de modifications a pu être proposé aux experts.

Nous avons ainsi déterminé :

1. Les règles jamais déclenchées. Certaines d'entre elles ont alors été supprimées, telles les règles :

"endométrite et acquisition_hôpital est en faveur STAPH <.2>",

"endométrite et infection urinaire est en faveur PENI_S <.6>".

D'autres règles ont été modifiées dans le sens d'une généralisation pour affaiblir les conditions de leur déclenchement. La règle,

"Asplénie et nourrisson est en faveur PENI_S <.5>" est devenue

"Asplénie et AGE <=15 est en faveur PENI_S <.5>".

2. Les redondances inutiles (selon les observations de la base de cas).

Examinons un exemple. Les experts avaient introduit les deux règles :

(R1) "Bpeumopathie (A) est en faveur PENI_S <.5>",

(R2) "Bpeumopathie (A) et BPCO (B) est en faveur PENI_S <.5>",

sachant bien qu'en présence de l'association "Bpneumopathie BPCO",

"PENI_S" se verrait affecter de la vraisemblance 0,75.

Cependant, les analyses bivariées ont montré que $VPP(A) = VPP(A \text{ et } B)$. L'association A et B ne semble donc pas apporter d'information supplémentaire par rapport à la donnée de A seul. Nous avons donc proposé la suppression de la règle (R2).

3. Proposer des suppressions de règles.

Considérons les quatre règles de la base de connaissances faisant intervenir le signe "acquisition chirurgie", règles confrontées aux données correspondantes du tableau T :

- (R1) acquisition chirurgie est en faveur STAPH <.3> ,
- (R2) acquisition chirurgie est en faveur AMPI_S <.3> ,
- (R3) acquisition chirurgie est en faveur BGN_VILLE <.2> ,
- (R4) acquisition chirurgie est en faveur BGN_HOSPI <.4> ,

	STAPH	PENI_S	AMPI_S	BGN_VILLE	SALMO	BGN_HOSP	AUTRE:
ACQ CHIR ¹ .							
Effectifs (53)	24	5	2	14	1	5	2
VPPi	45%	9%	4%	26%	2%	9%	3%
CVO_Pi	0,3	0	0	0	0	0,03	0
prévalences	22,5%	27,6%	6,5%	29%	5%	6%	3%

¹ Acquisition chirurgie

Le "BGN_VILLE" est assez fréquent lorsque la septicémie a été acquise en service de chirurgie, ce qui explique l'existence de la règle R3. Cependant, on constate, à partir des données de ce tableau, que la valeur prédictive positive du signe "acquisition chirurgie" pour le diagnostic de "BGN_VILLE" est proche de sa prévalence observée sur la base de cas. Nous avons donc proposé la suppression de cette règle R3. De la même façon, nous avons proposé la suppression des règles R2 et R4.

4. Proposer des modifications de coefficients de vraisemblance et proposer des règles.

A partir du tableau général T, on observe pour le signe "endocardite" :

- (R1) endocardite est en faveur STAPH <.3>,
 (R2) endocardite est en faveur PENI_S <.6>,

	STAPH	PENI_S	AMPI_S	BGN_VILLE	SALMO	BGN_HOSP	AUTRES
<u>ENDOCARDITE</u>							
Effectifs (76)	22	35	19	0	0	0	0
VPPi	30%	47%	26%	0	0	0	0
CVO_Pi	0,1	0,27	0,21	0	0	0	0
prévalences	22,5%	27,6%	6,5%	29%	5%	6%	3%

A partir des données de ce tableau, nous avons proposé de modifier les coefficients de vraisemblance des règles (R1) (0,1 au lieu de 0,3) et (R2) (0,3 au lieu de 0,6).

Nous avons également proposé d'introduire la règle :

- (R3) Endocardite est en faveur AMPI_S <.2>,

Nous avons enfin proposé d'introduire les règles négatives :

- (R3) Endocardite est en faveur NON BGN_VILLE <.9>,
 (R4) Endocardite est en faveur NON SALMONELLE <.9>,
 (R5) Endocardite est en faveur NON BGN_HOSPI <.9>,

L'ensemble des modifications suggérées par l'étape III figurent dans les tableaux de l'annexe II.

Ces propositions de modifications ont été soumises aux experts du GROUPE SES. 9 suppressions de règles ont été refusées tandis que 4 règles nouvelles, suggérées par notre étude, n'ont pas été jugées pertinentes par les experts.

Cependant, la plupart des modifications proposées ont été acceptées par les experts du GROUPE SES. C'est ainsi que :

- 53 règles ont été supprimées. Ces règles utilisaient, pour déduire un diagnostic donné D_i , des signes ou associations de signes dont les valeurs prédictives pour D_i étaient proches de la prévalence de D_i .

- 40 coefficients de vraisemblance ont été modifiés en utilisant la formule de SHORTLIFFE.

- 43 nouvelles règles ont été introduites. Certaines d'entre elles font intervenir des signes qui étaient déjà utilisés dans d'autres règles de la base de connaissances. D'autres nouvelles règles utilisent des signes qui n'avaient pas été jugés a priori pertinents par les experts pour le diagnostic des germes : "corticoïde", "radiothérapie", "pathologie colorectale"...

4.2 Evaluation de la base de connaissances DIAGSES3

La nouvelle base de connaissances prenant en compte l'ensemble des modifications validées par les Experts a été évaluée sur les 555 dossiers de notre sélection d'apprentissage.

Le **TABLEAU 4.II** donne les pourcentages d'identification des GBSV (Note 1 et Note 2) obtenus pour la nouvelle base de connaissances.

GBSV	EFFECTIFS (%)	EFFECTIFS NOTE 1 (%)	EFFECTIFS NOTE 2 (%)
STAPH	125 (22,5%)	82 (65,5%)	110 (88%)
PENI_S	153 (27,6%)	125 (81,5%)	149 (97%)
AMPI_S	36 (6,5%)	13 (36%)	31 (86%)
BGN_VILLE	164 (29,6%)	122 (74,5%)	145 (88%)
BGN_SALMONELLE	26 (4,7%)	16 (61,5%)	18 (69%)
BGN_HOSPITALIERS	34 (6,1%)	10 (30%)	27 (68%)
AUTRES GROUPES	17 (3%)	0	6 (35%)

TABLEAU 4.II

Ces résultats montrent que les performances du système sont très sensiblement améliorées : 66,5% des GBSV sont maintenant identifiés dès la première position (+10,6%) alors que 87,5% (+4,5%) de ces groupes sont identifiés dans l'une des trois premières positions.

On retrouve, pour le groupe BGN_VILLE, un pourcentage d'identification comparable à celui obtenu par la première base de connaissances. Les groupes GBSV sont tous mieux identifiés. En particulier, les "AMPI_S" et "BGN_SALMONELLE" sont maintenant correctement classés dès la première position.

Le tableau de fréquences (GBSV_REELS)x(PREMIERE CONCLUSION) reproduit ci-dessous permet d'étudier la qualité des réponses de DIAGSES3.

DIAGSES3 (Conclusion N*1) GBSV REELS	EFFECTIFS						
	STAPH	PENI_S	AMPI_S	BGN_VILLE	BGN SALMO.	BGN HOSPI.	AU'
STAPH	82	29	3	5	1	2	
PENI_S	15	125	6	2	0	2	
AMPI_S	1	16	13	5	0	0	
BGN_VILLE	17	10	1	122	6	4	
SALMONELLE	1	6	0	2	16	1	
BGN_HOSPI.	5	6	0	8	2	10	
AUTRES	6	4	0	2	2	3	

TABLEAU 4.III

Les réponses du système ont été jugées cohérentes par les experts. Les principaux problèmes qui subsistent, erreurs de discrimination entre les groupes "STAPH"/"PENI_S", "AMPI_S"/"PENI_S" et "BGN_HOSPITALIER"/"BGN_VILLE", correspondent à des GBSV qui peuvent être associés à des tableaux cliniques voisins.

Ces bons résultats montrent que les analyses bivariées peuvent apporter une aide à l'amélioration des performances de la base de connaissances. Les modifications de coefficients de vraisemblance, les suppressions et ajouts de règles, par référence aux fréquences des signes ou syndromes observés dans la base de cas, se sont révélés pertinents. La base de données nous a également permis de construire des règles faisant intervenir des items qui n'avaient pas été retenus par les experts dans la base de connaissances initiale.

Cependant, ces analyses bivariées ne prennent pas en compte l'effet associatif des signes ou syndromes. Dans la base de connaissances précédente, ces effets associatifs de signes sont, soit pris en compte directement dans la prémisse d'une règle (association jugée a priori pertinente par les experts), soit gérés par le moteur d'inférence grâce à la combinaison des coefficients de vraisemblance, moyennant les hypothèses d'indépendance des prémisses dont nous avons parlé au chapitre I.

Pour étudier dans quelle mesure des combinaisons linéaires des signes ou syndromes intervenant en prémisses de ces règles peuvent améliorer les performances diagnostiques, nous avons réalisé des études multifactorielles.

5 CONSTRUCTION DE DIAGSES4 PAR PRISE EN COMPTE D'ANALYSES MULTIFACTORIELLES

5.1 Méthodologie

Notre objectif était de construire des scores, résumant l'information contenue dans plusieurs règles, et ayant un bon pouvoir prédictif pour le diagnostic. Pour ce faire, nous avons réalisé des analyses factorielles discriminantes (AFD) et utilisé la ROC analyse. Nous savons que l'AFD permet de construire des scores pour différencier des classes définies a priori. La construction de ces scores nécessite par conséquent la définition de sous-but, qui doivent correspondre à une étape logique dans le processus de décision.

Notre étude n'a pas concerné les GBSV à faible prévalence, en raison précisément de leur trop faible fréquence dans notre sélection. Notre but est donc d'améliorer le diagnostic des six groupes : "STAPH", "PENI_S", "AMPI_S", "BGN_VILLE", "SALMONELLE" et "BGN_HOSPI".

Dans le cadre de notre application :

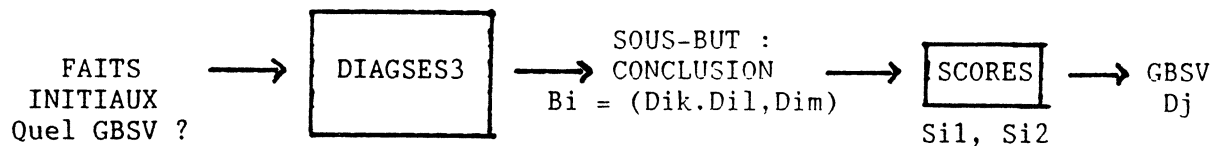
- a. il apparaît que ces groupes GBSV sont identifiés dans 90% des cas dans l'une des trois premières positions.
- b. Les experts ont analysé les 58 dossiers correspondant aux groupes GBSV non reconnus dans l'une des trois premières positions. 37 d'entre eux ont été jugés très difficiles : plus précisément, il n'était pas possible, selon les experts, de déterminer le diagnostic réel dans l'une des trois premières positions.

Le système expert est donc très performant, si on le juge sur ses trois premières réponses (90%). Pour l'identification en première position, ce pourcentage n'est plus que de 66%.

Il apparaît donc logique de faire confiance au système expert pour le diagnostic du GBSV dans l'une des trois premières positions et d'opérer ensuite une démarche de discrimination. Le système expert fournit, dans sa conclusion, un, deux, le plus souvent trois GBSV possibles. Ce dernier cas intéresse 329 dossiers sur les 555 cas de notre sélection. Notre étude a ainsi concerné l'amélioration des performances du système pour les cas, les plus nombreux, où trois réponses sont proposées.

Le problème peut alors être formalisé de la manière suivante :

- a. soit $D=(D_j, j=1,6)$ les six classes GBSV dont nous cherchons à améliorer le diagnostic.
- b. Une conclusion donnée B_i du système expert est constituée par un triplet de classes GBSV (D_{ik}, D_{il}, D_{im}) . Par exemple, pour un patient donné, le système fournit pour les trois premières réponses "PENI_S", "STAPH" et "BGN_VILLE".
- c. Chaque conclusion B_i peut être considérée comme un sous-but. Ces sous-buts B_i sont correctement atteints dans 90% des cas, dans la mesure où le GBSV réel appartient à B_i dans 90% des cas. Soit B l'ensemble des sous-buts. Il existe 20 sous-buts possibles : $CARD(B)=C(6,3)=20$.
- d. Pour améliorer la discrimination entre les éléments (D_{ik}, D_{il}, D_{im}) d'un sous-but donné B_i , nous avons réalisé une analyse factorielle discriminante sur les groupes $G_k, G_l,$ et G_m (G_k désigne l'ensemble des patients de la base de cas présentant le diagnostic D_k). Cette analyse permet de construire deux scores S_{i1} et S_{i2} , séparant le mieux possible (par combinaison linéaire) les trois groupes.
- e. 20 analyses factorielles discriminantes ont été réalisées pour construire vingt paires de scores discriminants.



5.2 Méthodologie adoptée pour la construction des scores

Les analyses ont concerné au total 501 dossiers : les 37 dossiers précédents pour lesquels les experts avaient jugé impossible la détermination du diagnostic ont été éliminés. Nous désignerons par base de cas, l'ensemble de ces 501 dossiers.

Pour chaque groupe Gik, Gil et Gim étudié, la méthodologie suivante a été adoptée :

- a. détermination, avec les experts, des variables participant à l'analyse.
- b. Réalisation d'une AFD sur l'ensemble de ces variables. Etude des résultats : qualité de la discrimination, variables les plus responsables de la discrimination, pourcentage de "bien classés", comparaison avec les performances du système expert. Cette étape permet de déterminer s'il est utile de poursuivre.
- c. Réalisation d'une analyse discriminante "pas à pas" pour isoler un sous-ensemble des variables les plus pertinentes pour séparer les groupes Gik, Gil et Gim.
- d. Réalisation d'une AFD sur ces variables. Etude des résultats, comme au b).
- e. Construction des scores à partir des composantes discriminantes. Etude de la sémantique de ces scores par comparaison avec les règles de la base de connaissances. Si nécessaire, détermination, par les techniques de la ROC analyse. des seuils discriminants. Calcul du pourcentage de "bien classés". de la sensibilité et spécificité associés à ces scores. Comparaison avec les réponses

du système expert.

- f. Etude de l'ensemble des résultats avec les experts : les scores sont-ils acceptables ?

Dans le prochain paragraphe, nous détaillerons l'une de ces analyses, correspondant à un sous-but particulier : "STAPH", "AMPI_S" et "PENI_S". Nous avons choisi d'exposer cette analyse pour les raisons suivantes :

- ces trois classes GBSV apparaissent proches dans les conclusions du système (TABLEAU 4.VI).
- ces trois GBSV sont fréquemment cités dans la conclusion du système (78 cas).
- Dans 68 cas cette conclusion est correcte (le GBSV réel est un "STAPH", un "PENI_S" ou un "AMPI_S"). Cependant, parmi ces dossiers, le pourcentage d'identification du GBSV réel en première position n'est que de 62% (26 erreurs).

Les nouvelles performances du système obtenues après la réalisation de l'ensemble de ces études seront ensuite présentées. On trouvera en annexe III, les graphiques correspondant aux vingt analyses effectuées.

5.3 Discrimination des groupes "STAPH", "AMPI S", "PENI-S"

L'étude a concerné 288 patients de la base de cas : 108 présentent un "STAPH", 148 un "PENI_S" et 32 un "AMPI_S". Nous désignerons par groupe "A-P-S" l'ensemble de ces 288 patients.

Avec les experts, 65 variables ont été sélectionnées pour participer à cette étude (TABLEAU 5.I). Ce sont les variables cliniques qui interviennent, isolément ou en association, dans les règles de la base de connaissances pour déduire l'un de ces trois groupes GBSV.

Abcès poumon	Endocardite gauche	Panophtalmie
Abcès cerveau	Endocardite droite	Plaie infectée
Acquisition Chir.	Epidurite	Polyarthrite
Acquisition hospi.	Erysypèle	Prothèse os
Acquisition Réa.	Ethylisme	Porte entrée cutanée
Acquisition Méd.	Encéphalite	Porte entrée cathéter
Acquisition Hémato.	Furoncle	Porte entrée urinaire
Acquisition Chir. dig.	Granulopénie	Porte entrée colique
Age	Hémopathie	Prostatite
Asplénie	Hémodialyse	Prothèse cardiaque
Ascite	Infection dent	Prothèse vasculaire
Arthrite	Infection cutanée	Pneumonie alvéolaire
Antibio. Antérieure	Infection ORL	Purpura
BPCO	Infection urinaire	Sonde urinaire
Bpneumopathie	Immunodépression	Suppuration rénale
Choc TI	Insuffisance rénale	Sida
Corticoïde	Lésion colorectale	Thrombophlébite suppu
Chimiothérapie	LLC	Toxicomane
Conscience anormale	Méningite	Traumatisme crânien
Derivation LCR	Ostéite	Tumeur solide
Diabète	Ostéomyélite	Valvulopathie
Dysglobulinémie	Pathologie colorectale	

TABLEAU 5.I

5.3.1 Analyse factorielle discriminante sur la totalité des variables

Une analyse factorielle discriminante a tout d'abord été réalisée sur l'ensemble des 65 variables.

La FIGURE 5.I représente la projection des individus sur l'unique plan factoriel discriminant. Ce graphique amène les remarques suivantes :

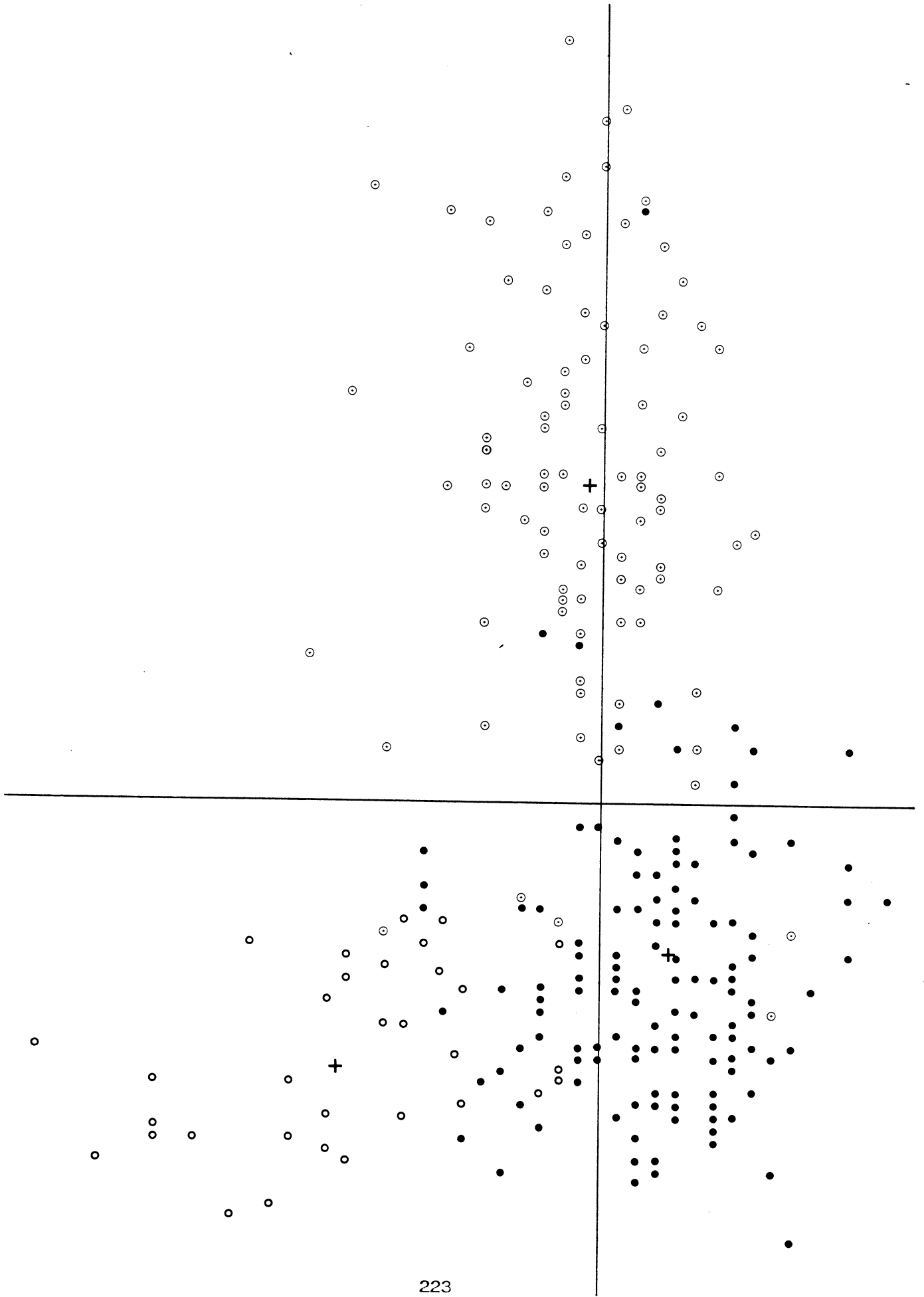


Figure 5.1 Analyse Factorielle Discriminante (65 variables, 288 individus)

CHAPITRE IV:- STRATEGIE D'AMELIORATION DES PERFORMANCES -

- a. Les groupes sont bien séparés par les composantes discriminantes. Les centres de gravité sont éloignés.
- b. La première composante discriminante (axe horizontal A1) sépare le groupe "STAPH" des groupes "PENI_S" et "AMPI_S". La deuxième composante discriminante (A2) sépare les deux groupes "PENI_S" et "AMPI_S".

Le TABLEAU 5.II donne les variables les plus fortement corrélées à chacune des deux composantes discriminantes :

VARIABLES	A1	VARIABLES	A2
Porte entrée cutanée	0,55	Valvulopathie	0,50
Arthrite	0,45	Infection urinaire	0,40
Acquisition Hospitalière	0,40	Endocardite gauche	0,35
Acquisition Chirurgie	0,35	Porte entrée urinaire	0,30
Plaie infectée	0,35	Pneumonie alvéolaire	-0,40
Spondylodiscite	0,30		

TABLEAU 5.II

Le TABLEAU 5.III donne le résultat du classement obtenu par la distance de Mahalanobis.

CLASSEMENT \ DIAGNOSTIC	DIAGNOSTIC			Pourcentage de "bien classés" : 90%
	STAPH	PENI_S	AMPI_S	
STAPH	97	4	0	
PENI_S	10	134	3	
AMPI_S	1	10	29	

TABLEAU 5.III

Il apparaît que les composantes discriminantes permettent de bien séparer les trois groupes étudiés. Ce bon résultat est à rapprocher des performances médiocres de DIAGSES3 pour le groupe A-P-S : 70 erreurs de discrimination pour ce groupe A-P-S (TABLEAU 4.VI). Cette comparaison doit cependant être relativisée dans la mesure où, dans le

cadre de notre étude et contrairement à DIAGSES3, nous supposons connue l'appartenance des patients au groupe A-P-S.

Le TABLEAU 5.II met en évidence les variables cliniques qui interviennent le plus pour séparer ces groupes. Les signes "Porte d'entrée cutanée", "Arthrite", "Plaie infectée", "Spondylodiscite" se retrouvent souvent chez les patients qui présentent une septicémie due au Staphylocoque. Les signes "Endocardite", "Spondylodiscite" et "Infection urinaire" sont en faveur d'une septicémie à "AMPI_S". La présence d'une Pneumonie alvéolaire s'oppose au contraire à une septicémie due au groupe "AMPI_S".

Nous reviendrons, dans la suite de notre exposé, sur ces caractéristiques moyennes des groupes "STAPH", "PENI-S" et "AMPI_S" mises en évidence par l'AFD.

Compte tenu de ces bons résultats, nous avons poursuivi notre étude en réalisant une analyse discriminante pas à pas.

5.3.2 Analyse discriminante pas à pas et deuxième AFD

Les deux composantes discriminantes du paragraphe précédent sont des combinaisons linéaires de l'ensemble des 65 variables. Certaines de ces variables peuvent contribuer très faiblement et donc de manière artificielle, à la séparation des groupes, d'autres variables peuvent être redondantes.

L'objectif de l'analyse discriminante pas à pas est d'extraire, des 65 variables de départ, un sous-ensemble de variables assurant la meilleure discrimination entre les trois groupes.

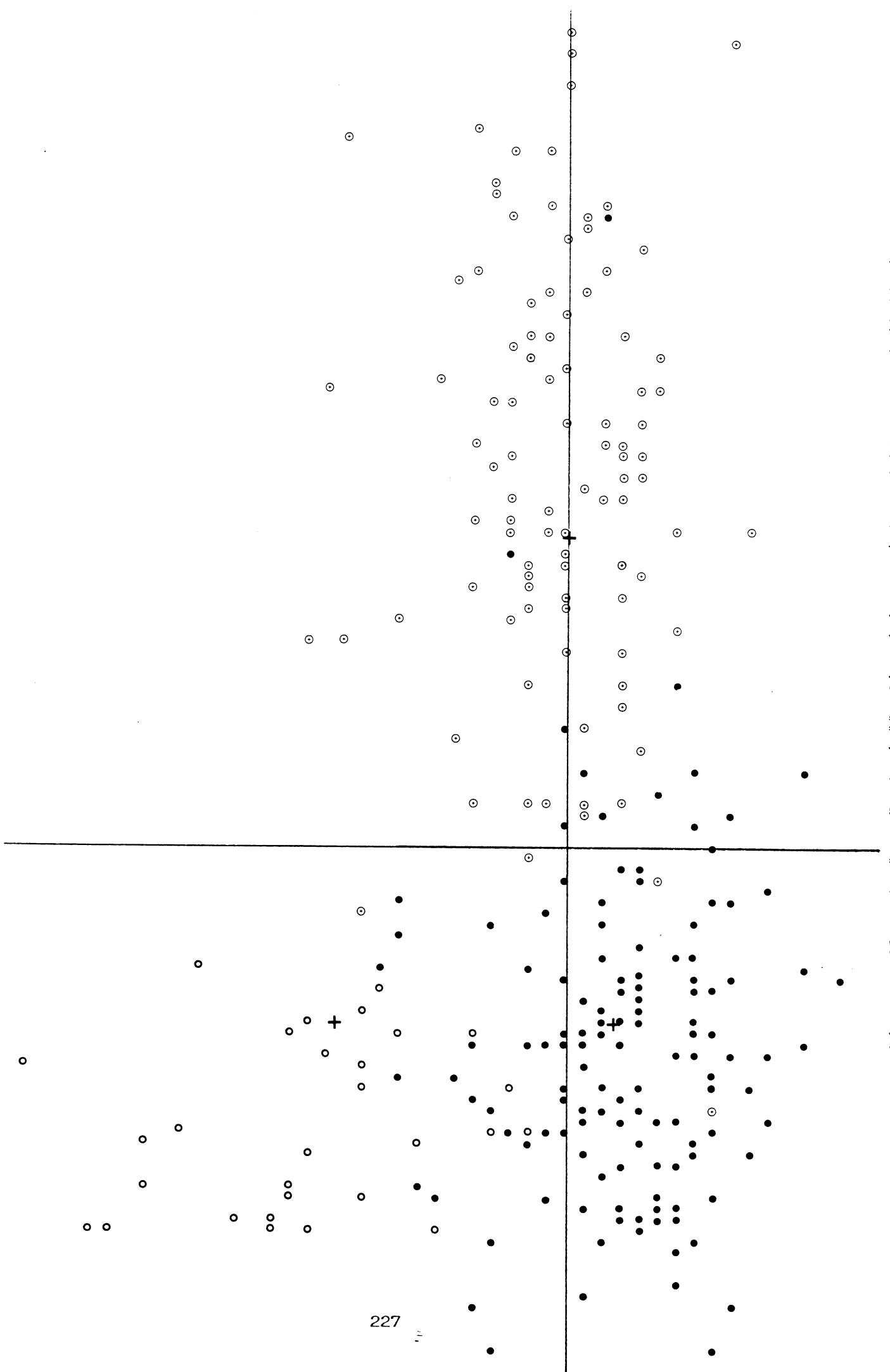
Un groupe de 25 variables a ainsi été déterminé (TABLEAU 5.IV).

porte entrée cutanée	valvulopathie
spondylodiscite	arthrite
infection urinaire	prothèse os
pneumonie alvéolaire	ostéite
sida	conscience anormale
infection dent	endocardite
sonde urinaire	méningite
thrombophlébite suppurée	prothèse cardiaque
acquisition chirurgie	acquisition hematologie
insuffisance rénale	abcès cerveau
chimiothérapie	acquisition réanimation
acquisition chirurgie- digestive	LLC
	âge

TABLEAU 5.IV

Nous avons ensuite réalisé une AFD sur ce sous-groupe de variables. La FIGURE 5.II représente la projection des patients dans le plan discriminant. On remarque que la qualité de la séparation des groupes par les composantes discriminantes est semblable à celle que nous avons observée sur la FIGURE 5.I à partir de l'étude sur 65 variables.

Le TABLEAU 5.V donne les variables les plus fortement corrélées (respectivement les moins corrélées) à chacune des composantes discriminantes A1 et A2 (respectivement aux deux composantes discriminantes) :



227

Figura 5.11 Análise Fatorial do Discriminante 195 variáveis 988 indivíduos

VARIABLES LES <u>PLUS CORRELEES</u> (à A1 ou A2)			
VARIABLES	A1	VARIABLES	A2
Pneumonie alvéolaire	-0,35	Pneumonie alvéolaire	-0,40
Endocardite gauche	-0,35	Endocardite gauche	0,50
Porte entrée cutanée	0,60	Valvulopathie	0,50
Arthrite	0,40	Infection urinaire	0,40
Ostéite	0,30		
Méningite	-0,30		
Acquisition chirurgie	0,40		
Spondylodiscite	0,40		

VARIABLES LES <u>MOINS CORRELEES</u> (à A1 et A2)			
	A1		A2
LLC	0,02		-0,03
Acquisition chirurgie digestive	0,06		0,06
Acquisition hématologie	-0,06		-0,05
Prothèse cardiaque	-0,01		-0,005

TABLEAU 5.V

Les données de ce tableau seront interprétées dans le paragraphe suivant.

5.3.3 Construction des scores de discrimination

Les composantes discriminantes (A1 et A2) fournies par l'AFD sont les deux combinaisons linéaires des variables prises en compte (25), dont les valeurs sont les plus différentes possibles pour les trois groupes étudiés ("STAPH", "PENI_S" et "AMPI_S").

A partir des coefficients de ces deux combinaisons linéaires, il est possible de construire deux scores, S1 et S2, destinés à différencier les trois groupes étudiés.

Cependant, il n'est généralement pas nécessaire de retenir l'ensemble des variables qui interviennent dans les combinaisons linéaires A1 et A2 [ROBERT 86] : l'étude des corrélations entre les variables et les composantes discriminantes permet en effet de déterminer celles qui contribuent très faiblement à la discrimination entre les groupes.

Dans le cadre de notre étude, nous avons éliminé quatre variables : "LLC", "acquisition chirurgie digestive", acquisition hématologie" et "prothèse cardiaque". Ces quatre variables sont très faiblement corrélées aux deux composantes discriminantes (TABLEAU 5.V) et, par conséquent, interviennent très peu pour séparer les trois groupes. De plus, les règles de la base de connaissances qui utilisent ces signes en prémisse ont des coefficients de vraisemblance peu élevés.

Les scores S1 et S2 que nous avons construits pour discriminer les groupes "STAPH", "PENI_S" et "AMPI_S", font ainsi intervenir 21 variables. Les poids associés à ces scores sont les coefficients des combinaisons linéaires A1 et A2 pour ces 20 variables (TABLEAU 5.VI).

	S1	S2
porte entrée cutanée	1,5	0,1
valvulopathie	0,5	1,6
spondylodiscite	2,8	0,2
arthrite	1,6	0
infection urinaire	-0,5	3,8
prothèse os	0,1	2,7
pneumonie alvéolaire	-0,8	-0,9
ostéite	1,5	-0,2
sida	-0,5	6,5
âge	-0,01	0,004
conscience anormale	0	-0,8
infection dent	-0,9	-1,8
endocardite	-0,7	1
sonde urinaire	2,4	-0,3
méningite	-0,7	0,6
thrombophlébite suppurée	2,2	0,3
acquisition chirurgie	2,1	-0,7
acquisition réanimation	2	0
insuffisance rénale	0,1	0,7
abcès cerveau	2	-1,8
chimiothérapie	1,9	0,3

TABLEAU 5.VI

Pour un patient donné, il est donc possible de calculer les valeurs des scores S1 et S2. Considérons par exemple un patient , âgé de 40 ans, présentant une septicémie à porte d'entrée cutanée, acquise en chirurgie. Les valeurs de S1 et S2 sont alors : $S1 = 3,2$ et $S2 = -0,45$. Pour ce patient, qui ne présente aucun des quatre signes éliminés pour la construction des scores S1 et S2 ("LLC"...), les valeurs de S1 et S2 sont égales à celles de A1 et A2. Si l'on se reporte au graphique de la FIGURE 5.II, ce patient se projette dans la zone occupée par le "STAPH".

Plus généralement, soit j un patient donné. Désignons par $S1(j)$ (respectivement $S2(j)$) la valeur de S1 (respectivement de S2) pour ce patient. En se reportant à la FIGURE 5.II, les scores S1 et S2 permettront alors d'affecter le patient j dans l'un des trois groupes étudiés de la manière suivante :

- si $S1(j)$ est élevée, alors le GBSV en cause est très probablement un "STAPH",
- si $S1(j)$ est faible et $S2(j)$ élevée, alors il s'agit très probablement d'un "AMPI_S",
- si $S1(j)$ et $S2(j)$ sont faibles, alors il s'agit très probablement d'un "PENI_S". Il est donc nécessaire :

- a) d'étudier si ces scores ont un sens clinique,
- b) de définir les seuils de décision (que signifie valeur élevée ? faible ?) et d'établir les paramètres statistiques caractérisant le pouvoir prédictif des scores : pourcentage de "bien classé", sensibilités, spécificités.

5.3.4 Etude de la sémantique des scores

Les scores sont fréquemment utilisés par les cliniciens (score de Diamond, score de Maccabe...). De tels scores de gravité ont une sémantique évidente. Notre démarche nous a permis de déterminer des scores à partir de l'AFD : ce sont des combinaisons linéaires des variables cliniques de départ, maximisant la dispersion entre les trois groupes étudiés. Ces scores ont donc été établis mathématiquement et à partir des données de la base de cas.

Il nous semble indispensable d'étudier la sémantique des scores déterminés par l'AFD, pour au moins deux raisons :

- les experts n'accepteront pas d'utiliser un score qui ressemblerait à une "boîte noire".
- Si les coefficients et les variables qui interviennent dans ces scores sont sans rapport avec les connaissances cliniques des experts, la discrimination obtenue risque d'être artificielle (trop finement

adaptée à la base de cas) et par conséquent instable (non adaptée à d'autres cas).

Pour étudier la sémantique d'un score, nous avons comparé les coefficients relatifs à chaque variable dans la combinaison linéaire (TABLEAU 5.VI), aux règles de la base de connaissances. Une telle comparaison doit cependant être réalisée avec prudence, dans la mesure où le coefficient attaché à une variable renseigne sur l'influence de cette variable conjointement à l'ensemble des autres variables de la combinaison linéaire [ROMEDER 73].

5.3.4.1 Principaux signes en faveur de "STAPH" (coefficients positifs de S1)

On retrouve parmi ces variables, celles qui sont fortement corrélées à la première composante discriminante (TABLEAU 5.VI).

Dans le TABLEAU 5.VII, nous avons résumé les résultats de la manière suivante :

- **colonne I, signes spécifiques** pour "STAPH" :

la présence d'un "X" indique que le signe intervient, dans la base de connaissances, en prémisse d'une seule règle pour déduire "STAPH" (parmi les règles du groupe A-P-S). De la même façon :

- **colonne II, signes non spécifiques mais prépondérants** pour le diagnostic de "STAPH" :

le signe est également utilisé en prémisse d'une règle pour déduire "PENI_S" ou "AMPI_S", mais le CV de la règle qui déduit "STAPH" est prépondérant.

- **Colonne III, contradictions** entre le score et les règles de

DIAGSES3 :

il existe une règle utilisant ce signe en prémisse pour déduire "PENI_S" ou "AMPI_S" avec un CV prépondérant.

	I	II	III
Porte entrée cutanée		X	
Spondylodiscite	X		
Arthrite		X	
Ostéite	X		
Sonde urinaire	X		
Thrombophlébite supp.		X	
Acquisition chirurgie	X		
Acquisition réanimation	X		
Abscès cerveau	X		

TABLEAU 5.VII

Les résultats de ce tableau montrent que pour le diagnostic de "STAPH", les coefficients et les variables du score S1 recourent les règles de DIAGSES3.

5.3.4.2 Sémantique des scores pour le diagnostic des classes "PENI_S" et "AMPI_S"

les scores apparaissent comparables aux règles de DIAGSES3. Ainsi, par exemple :

- les signes "pneumonie alvéolaire" ou "infection dent" sont en faveur de "PENI_S", dans les scores S1 et S2, comme dans la base de connaissances :

"Pneumonie alvéolaire est en faveur non AMPI_S <.9>"

"Pneumonie alvéolaire est en faveur PENI_S <.7>"

"Infection dent est en faveur PENI_S <.8>".

- Parmi les septicémies du groupe A-P-S, la présence d'une infection urinaire est en faveur d'un "AMPI_S". On retrouve une règle, présente dans la base de connaissances originelle et qui avait été supprimée à

l'étape III, en raison de la fréquence prépondérante de la classe "BGN_VILLE" parmi les septicémies à infection urinaire.

5.3.5 Détermination des seuils décisionnels

Un score S peut être assimilé à un examen clinique conduisant à une résultat quantitatif. Ce score est destiné à séparer deux populations, la première que nous appellerons "M" et la seconde "NM", en fonction d'un seuil "s" à déterminer :

	M	NM
S > s	VP ¹	FP
S ≤ s	FN ²	VN

$$\text{Sensibilité} = se = \frac{VP}{VP + FN} ; \quad \text{Spécificité} = sp = \frac{VN}{FP + VN}$$

¹ vrais positifs; ² faux négatifs.

S'il existe un seuil s_0 tel que $se = sp = 1$, les deux populations seront parfaitement séparées par les règles :

$S_1 > s_0$ alors "M" et $S_1 \leq s_0$ alors "NM".

En général, ce seuil théorique n'existe pas et le problème posé est celui de la détermination du "meilleur" seuil pour la discrimination entre les deux populations étudiées : déterminer s_0 tel que $FN + FP$ soit minimum.

Cependant, pour différentes valeurs du seuil s , les quantités se et sp (et donc FN et FP) varient en sens inverses. La ROC analyse permet de visualiser le pouvoir prédictif d'un score selon différents seuils et de choisir un seuil optimal. c'est-à-dire adapté au problème. Cette courbe est en effet construite en reportant, pour différentes valeurs du seuil, le point M de coordonnées $(1-sp, se)$

(FIGURE 5.III). Selon les types de problèmes, on peut considérer qu'il est préférable que le score S soit très sensible (minimiser FN : $M1$) ou au contraire très spécifique (minimiser FP : $M2$), et choisir un seuil en conséquence. Dans le cadre de notre problème de discrimination, nous avons choisi le seuil correspondant point (M_0) le plus proche du point idéal de coordonnées $(0,1)$.

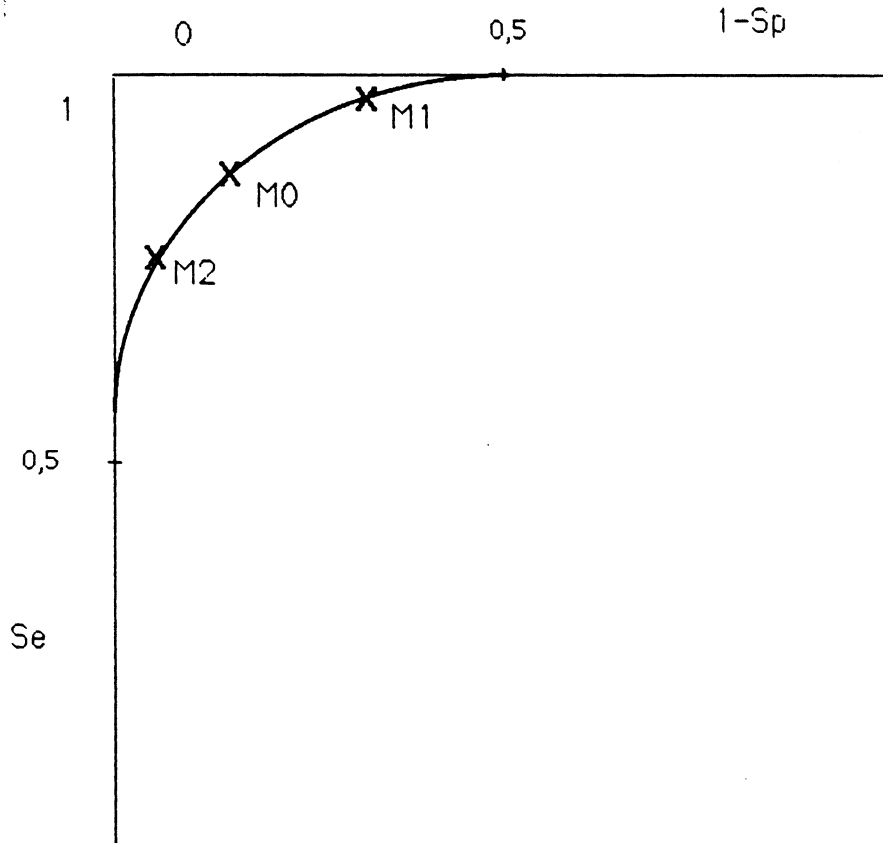


FIGURE 5.III

La FIGURE 5.II montre que la discrimination entre les trois groupes GBSV étudiés est décomposable en deux étapes : le score $S1$ est destiné séparer les groupes "STAPH" et "NON STAPH" et le score $S2$, dans la sous-population des "NON STAPH", les groupes "AMPI_S" et "PENI_S".

CHAPITRE IV:- STRATEGIE D'AMELIORATION DES PERFORMANCES -

Dans ces conditions, nous avons appliqué la méthode ROC pour déterminer deux seuils, "s1" et "s2" de telle façon que :

- a. au seuil "s1", le score S1 sépare le mieux possible les individus qui présentent un "STAPH" de ceux qui présentent un "PENI_S" ou un "AMPI_S". Nous avons choisi la valeur $s1 = 0,2$, qui correspondait au point le plus proche du point idéal (0,1) (FIGURE 5.IV).
- b. De la même façon, dans la sous-population des individus tels que " $S1 \leq s1$ ", le score S2 sépare le mieux possible les groupes "AMPI_S" et "PENI_S" pour $s2=1,2$.

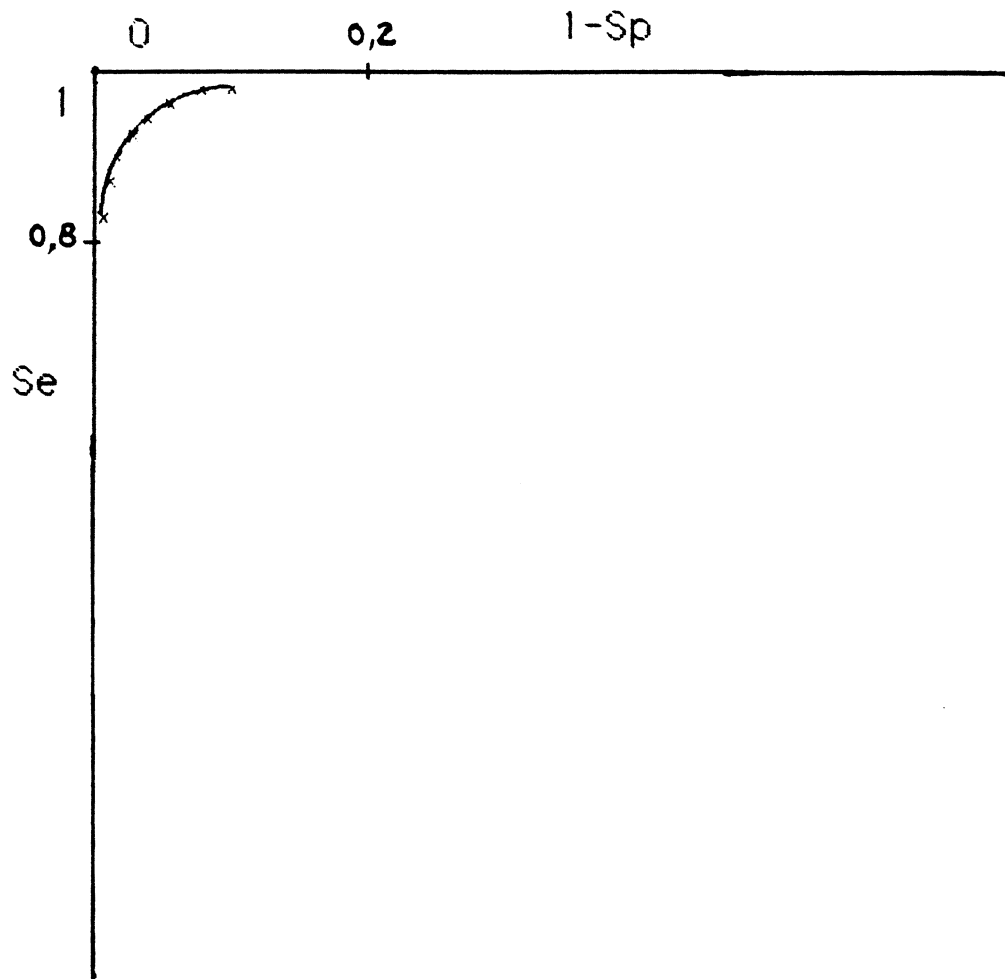


FIGURE 5.IV

Nous avons alors construit la variable qualitative C telle que pour l'individu j du groupe A-P-S :

$C(j) = \text{"STAPH"}$ si $S1(j) > 0,2$;

$C(j) = \text{"AMPI_S"}$ si $S1(j) \leq 0,2$ et $S2(j) > 1,2$;

$C(j) = \text{"PENI_S"}$ si $S1(j) \leq 0,2$ et $S2(j) \leq 1,2$;

5.3.6 Résultats

Le TABLEAU 5.VIII (C x Diagnostic réel) donne les résultats obtenus, à partir des scores S1 et S2, pour la discrimination au sein du groupe A-P-S entre les septicémies dues au "STAPH", celles dues au "PENI_S" et celles dues à l'"AMPI_S".

DIAGNOSTIC REEL CLASSEMENT PAR C	STAPH	PENI_S	AMPI_S
STAPH	98	9	0
PENI_S	9	138	5
AMPI_S	1	7	27

TABLEAU 5.VIII

Le pourcentage de "bien classés" est de 89%. Nous obtenons donc, avec 21 variables, un pourcentage de "bien classés" équivalent à celui obtenu à partir de la distance de Mahalanobis en tenant compte de l'ensemble des variables. Remarquons que cette distance n'est pas adaptée lorsque les groupes ont des formes spécifiques [SAPORTA 78]. C'est le cas ici, tout au moins pour le groupe "STAPH" qui est très "allongé".

Les scores que nous avons construits permettent donc de différencier les trois GBSV "STAPH", "PENI_S" et "AMPI_S". Ces scores apparaissent par ailleurs sémantiquement corrects.

Rappelons que 68 dossiers sont concernés par cette étude : Les trois premières réponses de DIAGSES3 pour ces dossiers sont "STAPH", "PENI_S" et "AMPI_S". Le TABLEAU 5.IX permet alors de comparer, sur ces 68 dossiers, les performances de DIAGSES4 pour l'identification du

GBSV en première position à celles de DIAGSES3.

		<u>DIAGSES3</u>					<u>DIAGSES4</u>		
DIAGNOSTIC REEL DIAGSES3		STAPH	PENI_S	AMPI_S	DIAGNOSTIC REEL SCORES (C)		STAPH	PENI_S	AMPI_S
		STAPH	13	2			0	STAPH	20
PENI_S	10	20	9	PENI_S	4	22	3		
AMPI_S	2	3	9	AMPI_S	1	3	15		
% de "bien classés" = 62%				% de "bien classés" = 84%					

TABLEAU 5.IX

Ces bons résultats montrent que les scores S1 et S2, combinaisons linéaires d'un sous-ensemble des signes pertinents intervenant en prémisses des règles du groupe A-P-S, permettent d'améliorer très sensiblement les performances du système pour ces 68 dossiers.

Les principales erreurs de classement de DIAGSES3, pour l'identification du GBSV en première position, se situent précisément au niveau de la discrimination entre les trois groupes "STAPH", "PENI_S" et "AMPI_S" (TABLEAU 4.VI). On pourrait donc envisager, dans une étape ultérieure, le groupement de ces trois GBSV en un sous-but et l'intégration de ces scores dans la base de connaissances sous forme de règles de production [ROBERT 86].

5.4 Résultats globaux

Une démarche équivalente a été appliquée pour les 19 autres conclusions possibles du système expert. On trouve en annexe III les graphiques correspondant à ces 19 analyses factorielles discriminantes. Ces graphiques montrent que les groupes GBSV sont

bien séparés par les composantes discriminantes.

Pour certaines de ces AFD, les résultats sont remarquables : les groupes sont parfaitement séparés par les composantes discriminantes. C'est le cas, par exemple, pour l'AFD sur les groupes "AMPI_S", "SALMONELLE" et "BGN_HOSPITALIER" à partir de laquelle on obtient 100% de "bien classés" (annexe III FIGURE 10). Sur le graphique correspondant, ces trois groupes GBSV apparaissent de plus très éloignés. Les résultats de cette AFD montrent que les tableaux cliniques associés à ces trois septicémies sont très différents. Cependant, ces trois GBSV n'apparaissent jamais simultanément dans les trois premières réponses du système expert. Il en est de même pour les autres groupes GBSV qui apparaissent très différents sur les graphiques de l'annexe III : FIGURE 2 et FIGURE 11, par exemple. Ce résultat démontre la cohérence des réponses de DIAGSES3.

Rappelons que les faits initiaux que l'on peut fournir au système expert sont des données de l'examen clinique et de la biologie standard. Généralement une dizaine d'items suffisent pour décrire le tableau clinique présenté par le patient.

Pour une conclusion donnée du système expert, $S_{Bi} = (Dik, Dil, Dim)$, les deux composantes discriminantes de la première AFD font intervenir l'ensemble des signes utilisés en prémisses des règles pour le diagnostic de "Dik", "Dil" et "Dim". Le calcul de scores associés à ces combinaisons linéaires est possible à partir d'une base de données, mais irréaliste en utilisation clinique réelle. De plus, et nos analyses le prouvent, un nombre important de ces variables contribuent très faiblement, et donc de manière artificielle, à la discrimination.

CHAPITRE IV:- STRATEGIE D'AMELIORATION DES PERFORMANCES -

Pour chaque sous-but SBi, nous avons donc déterminé deux scores S1i et S2i, calculés à partir d'un petit nombre de variables cliniques pertinentes, qui permettent d'améliorer le diagnostic différentiel entre les GBSV "Dik", "Dil" et "Dim" du sous-but SBi.

Pour construire ces scores, nous avons extrait un sous-ensemble des variables les plus pertinentes pour la discrimination entre ces diagnostics. Ainsi pour déterminer son diagnostic, le système DIAGSES4 calcule deux scores sur la base d'un maximum de 25 signes cliniques (TABLEAU 5.X). Il est donc possible de demander à l'utilisateur, si nécessaire, les informations complémentaires qui permettront le calcul de ces scores prédictifs (FIGURE 5.IV).

CONCLUSIONS	NOMBRE DE VARIABLES (pour le calcul des scores)
STAPH / PENI_S / BGN_VILLE	22
STAPH / PENI_S / SALMO.	18
STAPH / PENI_S / BGN_HOSPI.	25
STAPH / AMPI_S / BGN_VILLE	25
STAPH / AMPI_S / SALMO.	22
STAPH / AMPI_S / BGN_HOSPI.	25
STAPH / BGN_VILLE / SALMO.	24
STAPH / BGN_VILLE / BGN_HOSPI.	23
STAPH / SALMO. / BGN_HOSPI.	24
PENI_S / AMPI_S / BGN_VILLE	21
PENI_S / AMPI_S / SALMO.	19
PENI_S / AMPI_S / BGN_HOSPI.	25
PENI_S / BGN_VILLE / SALMO.	19
PENI_S / BGN_VILLE / BGN_HOSPI.	22
PENI_S / SALMO. / BGN_HOSPI.	24
AMPI_S / BGN_VILLE / SALMO.	18
AMPI_S / BGN_VILLE / BGN_HOSPI.	24
AMPI_S / SALMO. / BGN_HOSPI.	17
BGN_VILLE / SALMO. / BGN_HOSPI.	21

TABLEAU 5.X

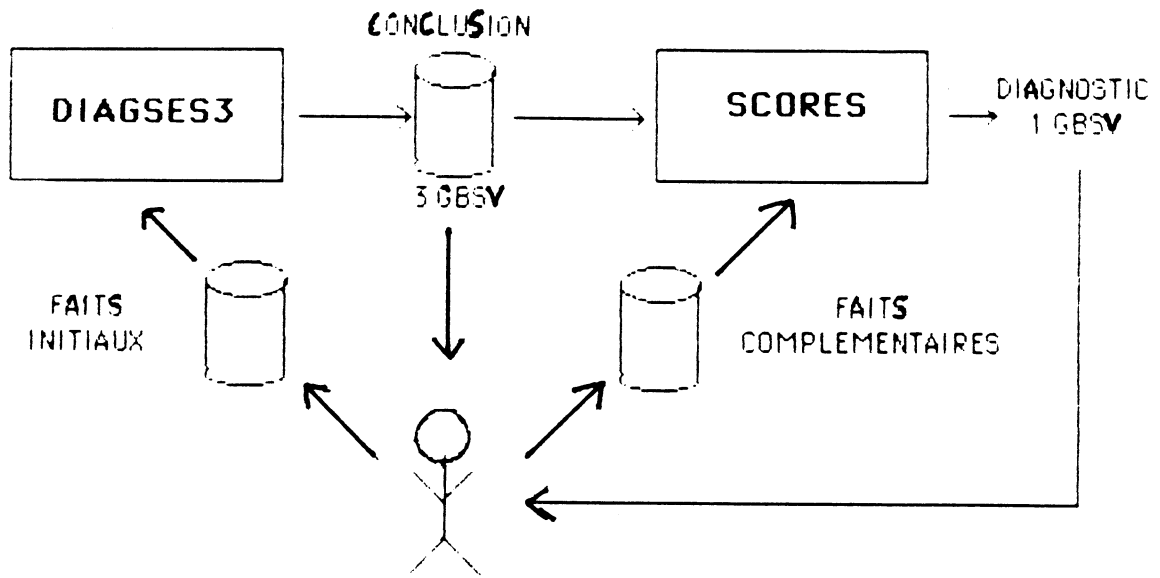


FIGURE 5.IV

Rappelons que 329 dossiers étaient concernés par les scores de discrimination. Sur ces 329 dossiers, les performances de DIAGSES3 et DIAGSES4 sont les suivantes:

RESULTATS SUR LES 329 DOSSIERS CONCERNES PAR LES SCORES:

	DIAGSES3	DIAGSES4
Identification en première position (%)	68%	84%

Le Tableau 5.XI donne les résultats obtenus par DIAGSES4 sur la totalité dossiers de la base de cas (555).

GROUPES GBSV	RAPPEL DIAGSES3	DIAGSES4
STAPH	65,6%	73%
PENI_S	81,5%	85%
AMPI_S	36%	58%
BGN_VILLE	74,5%	86%
SALMONELLE	61,5%	65%
BGN_HOSPITALIER	30%	53%
AUTRES	0	0
TOTAL (% "bien classés")	66,6%	75%

TABLEAU 5.XI

Les performances du système sont donc augmentées de près de 9%. Ces bons résultats montrent que des combinaisons linéaires de variables pertinentes permettent d'améliorer la discrimination entre les buts du système expert, les classes GBSV.

6 CONCLUSION

La complexité liée à l'évaluation des performances d'un système expert a été soulignée au cours du chapitre I. Nous avons utilisé, pour mesurer les améliorations successives apportées à la base de connaissances, le critère bien classé - mal classé (en première position ou dans l'une des trois premières positions).

Nous savons que cet indicateur ne permet pas, à lui seul, d'évaluer les performances du système expert. Dans notre application, par exemple, on obtiendrait 57% de "bien classés" dans l'une des trois premières positions si l'on choisissait de restreindre l'ensemble des conclusions fournies par le système aux trois germes les plus

fréquents: Ecoli - Staphylocoque - Pneumocoque. Un tel critère est donc en particulier insuffisant pour comparer, dans le cadre d'une même application médicale, les performances de deux systèmes d'aide à la décision. Il serait alors nécessaire de prendre en compte le nombre et la nature des conclusions possibles et leurs prévalences.

Cependant, cet indicateur nous a permis de montrer la pertinence de notre démarche en comparant ce qui est comparable, c'est-à-dire les valeurs de ce critère avant et après chaque modification. Nous avons en outre fourni aux experts le tableau (conclusion du système X conclusion réelle). L'analyse de ce tableau leur a permis d'apprécier la qualité des réponses du système. Ainsi, les modifications apportées à la première base de connaissances GBSV lors de l'étape II n'ont pas fait apparaître d'amélioration au sens du pourcentage de "bien classés"; pourtant, les réponses du système ont été jugées plus fiables par les experts, parce que les classes de germe provoquant des infections sévères (BGN hospitalier...) étaient mieux identifiées.

Notre démarche s'appuie donc sur un contrôle par les experts du processus de consolidation de la base de connaissances. Comme nous l'avons relevé, cette participation active des experts permet de juger de la qualité réelle des réponses du système amélioré en terme de pourcentage de "bien classés". Elle nous apparaît en outre nécessaire pour les raisons principales suivantes:

- a. notre méthodologie permet une adaptation de la base de connaissances aux cas réels. Mais une démarche purement statistique ne prend pas en compte la notion d'utilité diagnostique (nécessité de citer le *Listeria* en cas de grossesse par exemple).

- b. L'approche statistique que nous avons décrite ne prend pas en compte les connaissances "profondes" telles que les relations de causalité.
- c. Notre méthodologie ne permet pas de résoudre les problèmes posés par les conclusions très rarement observées.

Pour automatiser le processus, il conviendrait de construire une couche d'apprentissage permettant d'apprendre les changements, de surveiller les cas rares, de détecter les comportements d'urgence.

La stratégie d'amélioration que nous venons d'exposer, nous a donc permis, en quatre étapes successives, d'améliorer de plus de 26% les performances diagnostiques du système expert SE-SES sur les 555 dossiers d'apprentissage de la base de cas :

BASE DE CONNAISSANCES ORIGINELLE (DIAGSES) : 48,5% (germes)
BASE DE CONNAISSANCES DIAGSES3 ET SCORES : 75% (GBSV)

Cette stratégie d'amélioration repose sur des analyses (évaluations - études statistiques) effectuées sur une base d'apprentissage. Elle doit être complétée par une étude de la stabilité des résultats obtenus. Les modifications successives apportées à la base de connaissances ont ainsi été contrôlées sur une nouvelle base de cas : les 290 dossiers de septicémie de la base de données BD_SES de l'année 1987. Ces dossiers n'avaient participé à aucune des analyses statistiques et aucune évaluation du système expert réalisées au cours de ce travail. Les résultats obtenus sur cet échantillon test sont détaillés dans le tableau récapitulatif de notre conclusion générale. Ce tableau montre que les performances de chacune des bases de connaissances successives ne sont pas

significativement différentes (au seuil 5%) selon les deux échantillons, apprentissage - test.

CHAPITRE V : APPLICATION DE LA METHODOLOGIE DES RESEAUX BAYESIENS

CHAPITRE V : APPLICATION DE LA METHODOLOGIE DES RESEAUX BAYESIENS

Nous avons appliqué au problème du diagnostic des classes GBSV, la méthodologie des réseaux bayésiens proposée par PEARL [PEARL 86 87 88]. Nous disposons en effet de toute la connaissance nécessaire à la réalisation d'un tel modèle de représentation des connaissances : les signes, syndromes et diagnostics GBSV de la base de connaissances DIAGSES3, constituent l'ensemble des noeuds du réseau ; les règles de la base de connaissances expriment les liens existant entre les noeuds du réseau ; ces liens ont été explicités par les études statistiques précédemment réalisées.

On peut donc construire les noeuds et les arcs du réseau et l'on dispose, grâce aux données statistiques déduites de la base de cas, des poids associés à ces arcs.

Notre objectif était de comparer les performances d'un tel modèle de représentation des connaissances à celles de la base de règles DIAGSES3 [BEUSCART 89], [DELEU 88a], [DELEU 88b]. Le modèle des réseaux bayésiens permet en effet d'explicitier les dépendances et les indépendances entre les items du domaine, et permet également une représentation numérique exacte et cohérente, en termes de probabilités conditionnelles, de l'incertitude du raisonnement.

1 PRESENTATION DU RESEAU UTILISE

1.1 Introduction

Les réseaux probabilistes (FIGURE 1.I) utilisent des graphes dirigés acycliques. La direction de l'arc exprime la dépendance, au sens bayésien, de l'item extrémité finale vis-à-vis de celui extrémité

initiale. Le graphe ne possède pas de cycles : dans le domaine médical, la présence d'un diagnostic coïncide avec l'apparition de certains symptômes dont il est la cause et non la conséquence.

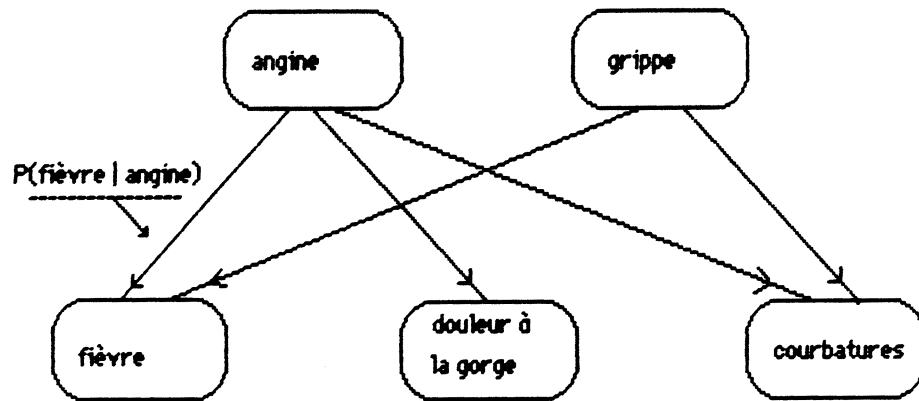


FIGURE 1.I

Considérons la suite ordonnée X_1, X_2, \dots, X_N , constituant l'ensemble des nœuds du réseau. Alors à chaque X_i est associé S_i , l'ensemble des prédécesseurs directs de X_i , tel que :

$$P(X_i/S_i) = P(X_i/X_1, X_2, \dots, X_{i-1}).$$

Chaque poids associé à un arc est donc la probabilité conditionnelle de la réalisation d'une proposition " $X_i=x_i$ ", étant donné l'état des nœuds de S_i , ses **prédécesseurs directs**. Pour assurer la cohérence avec la théorie des probabilités, il faut garantir la consistance de chacun des poids et que leur ensemble détermine complètement le savoir.

Comme le souligne [PEARL 86], les règles de production définissent une compréhension intuitive des contraintes majeures du domaine d'expertise, mais ne permettent pas une représentation numérique exacte de $P(X_1, X_2, \dots, X_N)$. Il existe alors un fossé entre la

représentation globale et l'estimation parcellaire des coefficients de vraisemblance.

1.2 Fonctionnement

Soit W , l'ensemble des noeuds du réseau. Désignons par e , le sous-ensemble des noeuds instanciés au début du fonctionnement. Partant de cet ensemble e , poser un diagnostic consiste à rechercher la meilleure instantiation w^* des noeuds de W , telle que w^* soit la meilleure explication de e , c'est-à-dire la plus probable [PEARL 87] :

$$P(w^*/e) = \max_W P(w/e).$$

Cette évaluation est réalisée localement sur chaque noeud, en prenant en compte simplement l'état de ses voisins directs par le biais de messages. On calcule pour chaque noeud X , son "BELIEF" (on pose $W_x = W - X$) :

$$Bel^*(X) = \max_{W'_x} \frac{P(X=vrai, w'_x | e)}{P(X=faux, w'_x | e)}$$

Nous allons établir des résultats pour $P(X=x, W'_x | e)$. Considérons la Figure 1.II, représentant un fragment de réseau bayésien, illustrant le découpage en sous graphes et le passage de messages entre les noeuds.

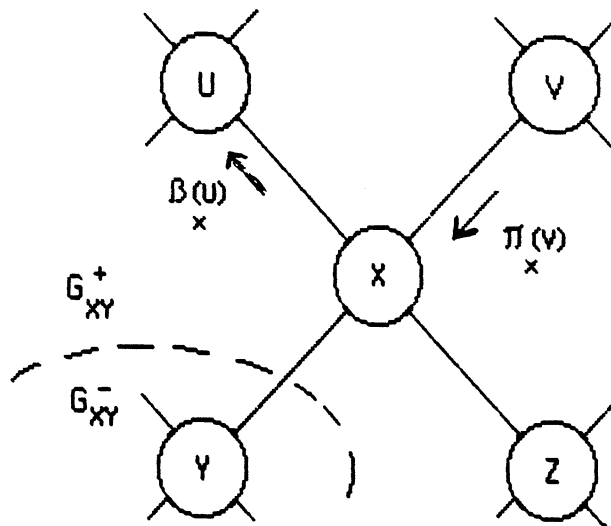


FIGURE 1.II

Le lien $X \longrightarrow Y$ divise le graphe en deux : G_{xy}^+ et G_{xy}^- , sous-graphes auxquels on associe les sous-ensembles W_{xy}^+ et W_{xy}^- . On obtient alors une partition de W :

$$W = W'_X \cup \{X\} \quad \text{où} \quad W'_X = W_{-XY} \cup W_{-XZ} \cup W_{+UX} \cup W_{+VX} \quad \text{et} \quad W_{+UX} = \{U\} \cup W_{+U'X}$$

On en déduit :

$$P(X=x, W'_X | e) = P(W_{-XY} | X=x) \cdot P(W_{-XZ} | X=x) \cdot P(X=x | U, V) \cdot P(U, W_{+U'X}) \cdot P(V, W_{+V'X}) \quad (1)$$

$$\text{On pose} \quad \beta_Y(X=x) = P(W_{-XY} | X=x) \quad \text{et} \quad \pi_X(U=u) = P(U=u, W_{+U'X})$$

$$\text{alors} \quad P(X=x, W'_X | e) = \beta_Y(X=x) \cdot \beta_Z(X=x) \cdot P(X=x | U, V) \cdot \pi_X(U) \cdot \pi_X(V) \quad (2)$$

L'équation (2) montre comment le calcul de $P(X=x, W'_X | e)$ est réalisé localement : le noeuds fils envoient les β -messages et les noeuds pères les π -messages. Au noeud X est associée une table contenant les valeurs $P(X=x | U, V)$.

Un noeud est activé par dépôt de messages β et/ou π ; les noeuds calculent leur probabilité de réalisation $P(X=x, W'_X | e)$, puis déposent à leur tour des messages. Le réseau s'arrête lorsqu'il n'existe plus de messages échangés.

L'algorithme détaillé peut être trouvé dans [PEARL 86]. Le calcul effectué au niveau de chaque noeud est $bel^*(X)$:

$$\text{si on note } \beta^*(X) = \max_Y \frac{P(W_{XY}^- | X=\text{vrai})}{P(W_{XY}^- | X=\text{faux})} \quad (3)$$

alors on obtient pour le belief:

$$Bel^*(X) = \beta^*(X) \cdot \beta^*(X) \cdot \max_{u,v} \frac{P(X=\text{vrai} | U, V)}{P(X=\text{faux} | U, V)} \quad (4)$$

2 APPLICATION AU DIAGNOSTIC DES GBSV

Comme nous l'avons vu au chapitre IV paragraphe 4.1, l'incertitude, dans le cadre du diagnostic des GBSV, n'affecte qu'une étape du raisonnement. La première étape permet de structurer les signes de plus bas niveau en syndromes.

Dans la base de données BD_SES, l'ensemble de ces signes et syndromes est enregistré pour chaque patient. Nous avons choisi, pour cette étude de comparaison, une représentation à deux niveaux, comme dans [PEARL 87] : le niveau inférieur est constitué de signes ou de syndromes, le niveau supérieur regroupe les dix classes GBSV.

Ces signes et syndromes sont ceux qui sont utilisés dans la base de connaissances DIAGSES3. Les arcs ont été construits à partir des règles de DIAGSES3. Par exemple, aux règles :

Endocardite est en faveur PENI_S <.3>.

Endocardite est en faveur AMPI_S <.2>.

nous avons associé deux arcs joignant le fait "endocardite" aux deux classes GBSV "PENI_S" et "AMPI_S" (FIGURE 2.I). Le poids porté sur un

arc reliant le symptôme "Si" au diagnostic GBSV "Gj" est la sensibilité du signe pour Gj : $P[Si/Gj]$. Ces sensibilités sont connues par les études statistiques réalisées sur la base de données (sélection des 555 dossiers des années 85-86).

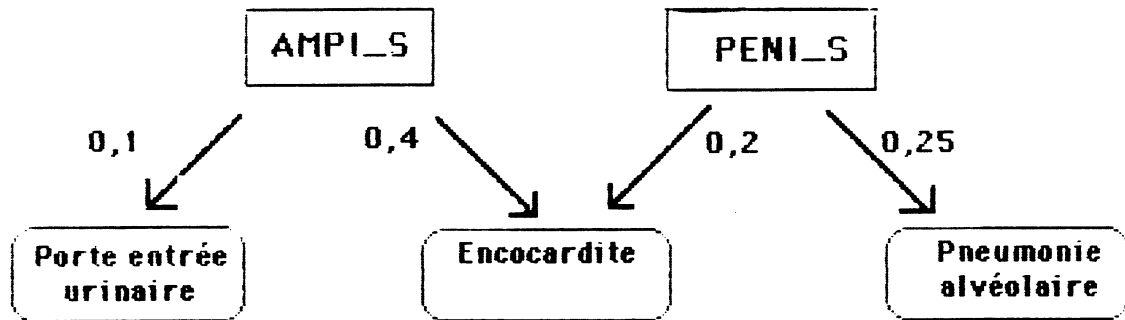


FIGURE 2.I

Le réseau appliqué à SES comporte 89 signes ou syndromes (niveau bas) et les dix classes GBSV (niveau haut).

Soit S un symptôme. Nous supposons comme dans PEARL [PEARL 87] que :

- S ne peut être observé que dans le cas où l'une de ses causes est présente (c'est-à-dire un GBSV qui lui est lié).
- La non observation de S ($S = \text{faux}$) en présence d'une classe GBSV G_i ($G_i = \text{vrai}$), est indépendante de sa non observation en présence d'une autre classe GBSV ($G_j = \text{vrai}$) :

$$P(S=\text{faux}, G_i=\text{vrai}, G_j=\text{vrai}) = P(S=\text{faux}, G_i=\text{vrai}) \cdot P(S=\text{faux}, G_j=\text{vrai}).$$

on démontre alors que:

$$P(S=\text{faux} | G_1, \dots, G_{10}) = \prod_{i \in I_T} P(S=\text{faux} | G_i=\text{vrai})$$

$$\text{où } I_T = \{ i / G_i = \text{vrai} \}$$

Dans le réseau, les noeuds "racine" ne possèdent bien sûr pas de pères. Dans les formules précédemment données, $P(X/U,V)$ devient alors $P(X)$. Dans notre application, X est une classe GBSV et $P(X)$ désigne donc la prévalence du GBSV X. Nous disposons d'une estimation de ces prévalences sur la base de cas (chapitre IV, paragraphe 2). Ces prévalences sont donc utilisées explicitement, lors de l'activation des noeuds GBSV.

3 RESULTATS

Nous avons tout d'abord soumis l'ensemble des 555 dossiers de la base de cas au réseau.

Le TABLEAU 3.I donne les résultats obtenus :

GROUPES GBSV (effectifs)	PERFORMANCES	
	RAPPEL DIAGSES3	RESEAU
STAPH (125)	65,6%	62%
PENI_S (153)	81,5%	75%
AMPI_S (36)	36%	53%
BGN_VILLE (164)	74,5%	77,5%
SALMONELLE (26)	61,5%	46%
BGN_HOSPITALIER (34)	30%	32%
AUTRES (17)	0	35% (6)
TOTAL % "bien classés":	66,6%	66%

TABLEAU 3.I

On remarque que ces résultats sont très proches, que ce soit globalement (pourcentage de "bien classés") ou par classe GBSV, de ceux obtenus avec la base de règles DIAGSES3.

Cependant, cette évaluation du réseau concerne les dossiers de la base de cas à partir desquels les poids des arcs (sensibilités des signes pour les GBSV) ont été déterminés. Il était par conséquent nécessaire de vérifier la stabilité de ces résultats sur une base de cas "test". Comme pour le système de règles de production, nous avons étudié le comportement de notre réseau sur les 290 dossiers de l'année 1986 (chapitre IV, paragraphe 6 et conclusion générale).

Le résultats que nous avons obtenus sur ces dossiers apparaissent dans le Tableau 3.II :

GROUPES GBSV (effectifs)	PERFORMANCES RESEAU
STAPH (62)	63%
PENI_S (77)	71,5%
AMPI_S (13)	36,5%
BGN_VILLE (82)	83%
SALMONELLE (20)	70%
BGN_HOSPITALIER (19)	21%
AUTRES (17)	0
TOTAL (% "bien classés")	64%

TABLEAU 3.II

Les performances sont donc stables : il n'existe pas de différences significatives (au seuil 5%) entre les fréquences d'identification pour chaque GBSV selon les deux échantillons.

Cette étude démontre l'intérêt des systèmes bayésiens pour résoudre les problèmes de diagnostic dans les domaines incertains pour lesquels il existe une connaissance statistique. Cette connaissance statistique est disponible lorsque le système expert est développé parallèlement à une base de données. Dans notre application, la construction du réseau a été facilitée par l'existence de la base de connaissances.

La base de règles d'un système expert est difficile à mettre au point et à maintenir, comme le démontre notre étude du chapitre IV. Ce n'est pas le cas du réseau bayésien, dont la maintenance peut être assurée grâce aux données statistiques de la base de cas.

CONCLUSION GENERALE

RESULTATS

Echantillon d'apprentissage : 555 dossiers (Années 1985-86)

Echantillon test : 290 dossiers (Année 1987)

DIAGSES : Base de connaissances originelle (25 germes)

DIAGSES1 : Base de connaissances pour l'identification GBSV (10 groupes)

DIAGSES2 : Modifications sur la base de l'évaluation

DIAGSES3 : Modifications sur la base des résultats statistiques
(Analyses bivariées).

DIAGSES4 : Prise en compte des scores des Analyses Factorielles
discriminantes

I - Echantillon d'apprentissage

Pourcentage de bien classés DIAGSES : 48,5%

	DIAGSES1	DIAGSES2	DIAGSES3	DIAGSES4
Staph	60%	63%	65,6%	73%
Peni-S	64,7%	73%	81,5%	85%
Ampi-S	0	3%	36%	58%
BGN-ville	76,8%	64%	74,5%	86%
BGN-Salmo	0	15%	61,5%	65%
BGN-HOSPI	15%	30%	30%	53%
AUTRES	0	0	0	0
% bien classés	55,3%	55,85%	66,5%	75%

II - Echantillon test

Pourcentage de bien classés DIAGSES : 48,3%

	DIAGSES1	DIAGSES2	DIAGSES3	DIAGSES4
Staph	67,7%	70%	74%	80%
Peni-S	66,2%	74%	70%	77%
Ampi-S	0	0	27%	55%
BGN-ville	82%	61%	79%	84%
BGN-Salmo	0	15%	85%	85%
BGN-Hospi	0	5%	16%	42%
Autres	1(1)	1	1	1
% bien classés	55,5%	54%	65%	72,5%

(1) : un candida a été identifié en première position.

CONCLUSION

Durant la dernière décennie, de très nombreux systèmes experts médicaux ont vu le jour. Depuis MYCIN, les systèmes se sont attachés à mieux formaliser la connaissance, à rendre compte de l'incertitude, à s'intégrer dans des mondes informatiques de plus en plus complexes (bases de données, télématique, carte à mémoire [BEUSCART 86], etc...). Néanmoins, leur utilisation reste limitée, voire marginale. Un des problèmes qui sous-tendent cette utilisation confidentielle réside sans doute dans le fait que ces systèmes d'aide à la décision médicale nécessitent une évaluation très sérieuse avant de pouvoir être diffusés.

Les médecins ne peuvent utiliser un nouveau produit pharmaceutique qu'après l'avoir dûment testé et qu'il ait fait l'objet de très nombreux contrôles et expertises divers. Il apparaît ainsi difficile aux médecins de ne pas soumettre aux mêmes exigences scientifiques et éthiques, les systèmes d'aide à la décision, que ce soit une aide à la décision diagnostique ou une aide à la prescription thérapeutique. WYATT a bien souligné le parallélisme entre essai thérapeutique et évaluation d'un système d'aide à la décision. Nous avons repris ce point de vue, dans notre travail, en soulignant néanmoins la difficulté que nous avons rencontrée à reproduire cette démarche, en particulier lorsque l'on évoque la possibilité d'examen comparatifs.

Notre propos est donc de proposer une stratégie qui permette d'évaluer la faisabilité, l'utilité et la pertinence d'un système d'aide à la décision médicale. Nous n'oublierons pas, dans tout ce qui suit, que le monde médical est un monde incertain où les

connaissances sont entachées à la fois d'imprécision (imprécision sémiologique, imprécision des mesures) et d'incertitude (est-on jamais certain d'un diagnostic ?).

Nous soulignerons également l'importance, dans tout ce qui suit, de pouvoir confronter l'expertise à la réalité du milieu dans lequel elle s'exerce. Il nous paraît ainsi tout à fait indispensable de pouvoir disposer, à brève échéance, d'une base de données (base de cas cliniques par exemple) qui serve de référence pour les études de faisabilité ou d'évaluation des systèmes d'aide à la décision. Dans cette base de cas cliniques, il est nécessaire de constituer deux sous-groupes : l'un d'évaluation-amélioration, l'autre servant de groupe test.

Dans certains cas, la faisabilité d'un système d'aide à la décision médicale ne laisse pas place au doute : le savoir est bien structuré, les processus de diagnostic aboutissant à une discrimination malade/non malade sont bien explicités par les experts (ce fut le cas, par exemple, du Système Expert Septicémies SES). Dans d'autres conditions cependant, le problème est moins bien structuré et l'on peut se demander si les faits recueillis, les signes enregistrés, sont suffisants, quoi qu'en pensent parfois les experts, pour établir les règles d'une base de connaissances et structurer un système expert. Dans ces cas, nous avons proposé une méthodologie d'analyse de données qui consiste à étudier les possibilités de discrimination malade/non malade, en fonction des paramètres qui devraient être utilisés dans le système expert médical. Nous avons ainsi démontré la non-faisabilité d'un système expert dans le cadre des douleurs coronariennes.

- CONCLUSION GENERALE -

Si le système expert est réalisé, il est nécessaire de l'évaluer. Pour cela, il faut disposer d'un critère de référence confirmant ou infirmant de façon non équivoque le résultat de l'expertise.

Dans le cadre de SES, nous disposons de ce "gold standard" sous la forme du résultat bactériologique, c'est-à-dire de l'identification du germe réellement à l'origine de l'infection. Il nous a donc été facile d'établir, à partir d'une base regroupant 845 cas, avec 107 items par cas, une méthodologie d'évaluation du système expert. Confrontant les résultats du système d'aide au diagnostic à l'observation bactériologique, il a été possible d'en analyser les performances.

Bien entendu, toute évaluation montre des erreurs, des carences et des imperfections. Nous avons proposé, au cours de notre travail, une méthodologie consistant à produire de nouvelles règles, à adapter les coefficients de pondération des règles en fonction des méthodes de calcul utilisées, pour mettre à jour les coefficients de certitude. Cette méthodologie s'est inspirée des propositions de SHORTLIFFE concernant les méthodes de calcul des coefficients de certitude dans MYCIN et de la théorie des probabilités conditionnelles, mais aussi de certains résultats d'analyses de données multifactorielles. Grâce à cette amélioration des CV, il a été possible d'obtenir une performance accrue du système.

Nous avons montré aussi que les systèmes d'aide au diagnostic pouvaient être optimisés en termes de performances, si on rajoutait, à l'intérieur du corps des règles ou suite à une première expertise, une connaissance de type probabiliste, sous forme de scores diagnostiques, ou de scores de discrimination. Ces scores ne sont pas étrangers au

- CONCLUSION GENERALE -

monde médical qui en use beaucoup, en particulier en réanimation, dans les unités de soins intensifs, en cardiologie ou en imagerie.

Leur intégration dans le cadre de systèmes d'aide à la décision nous apparaît comme très profitable, dans certaines conditions, à la puissance d'analyse et de discrimination du système. Le tableau de cette conclusion démontre la pertinence de cette démarche, qui nous a amenés en quatre étapes successives à une version très performante (75% de "bien classés" en première position du système expert).

Cette connaissance complexe et multiple, associant raisonnement logique, raisonnement incertain, voire scores de discrimination, nous a également amenés à évaluer la pertinence, dans le cadre d'une aide au diagnostic, d'une utilisation des réseaux bayésiens selon la méthodologie proposée par PEARL. Notons que la réalisation de ce réseau n'aurait pas été possible si les étapes préalables n'avaient pas été effectuées et si nous n'avions pas disposé d'une base de cas cliniques suffisante.

Mais cette démarche d'optimisation et d'amélioration d'une base de connaissances manipulant des données et des règles incertaines et imprécises n'a de sens que dans le cadre d'une démarche dialectique avec les cliniciens experts responsables du savoir médical et conscients de l'enjeu clinique en termes de stratégie diagnostique et thérapeutique. C'est parce que les cliniciens du Groupe SES ont constamment collaboré avec nous, que le système expert SES a pu obtenir l'état d'achèvement dans lequel il se trouve aujourd'hui, prêt à être diffusé par le réseau MINITEL.

ANNEXE I

Nom du patient.....
 Adresse.....
 N° Téléphone.....
 Médecin Traitant.....
 Adresse.....
 N° Téléphone.....
 Cardiologue.....
 Adresse.....
 N° Téléphone.....

Instructions : ce questionnaire doit être rempli par vous même.
 Les réponses se font par oui ou par non (entourez la réponse),
 il est possible de répondre plusieurs fois oui à la même
 question. Si toutefois vous ne compreniez pas certaines
 questions, n'y répondez pas ; un médecin vous aidera à remplir
 ce questionnaire.
 La précision de vos réponses nous permettra de mieux définir
 vos problèmes et nous l'espérons mieux vous traiter.

Date :

Age : Sexe : Poids : Taille :

1 - Avez-vous (ou avez-vous déjà eu) une douleur ou une gêne dans la poitrine ou dans les bras ? oui non

(si non, passer à la question 22)

2 - La douleur ou la gêne est (ou était) située :

a - au milieu de la poitrine ? (derrière le sternum)	oui	non
b - près du sein gauche ?	oui	non
c - près du sein droit ?	oui	non
d - d'un sein à l'autre ?	oui	non
e - dans le dos ?	oui	non
f - ailleurs dans le thorax ?	oui	non
g - dans le bras gauche ?	oui	non
h - dans le bras droit ?	oui	non
i - dans l'épaule gauche ?	oui	non
j - dans l'épaule droite ?	oui	non
k - dans le coude gauche ?	oui	non
l - dans le coude droit ?	oui	non
m - dans le poignet gauche ?	oui	non
n - dans le poignet droit ?	oui	non

3 - La douleur s'étendait-elle vers :

a - le cou ?	oui	non
b - les joues ?	oui	non
c - les mâchoires ou les dents ?	oui	non
d - le bras gauche ?	oui	non
e - le bras droit ?	oui	non
f - l'estomac ?	oui	non

4 - Est-ce que la douleur est (ou était) produite par ?

a - la marche ?	oui	non
-----------------	-----	-----

- | | | |
|--|-----|-----|
| b - monter les escaliers ? | oui | non |
| c - un rapport sexuel ? | oui | non |
| d - une émotion ou une contrariété ? | oui | non |
| e - un repas ? | oui | non |
| f - une marche après un repas ? | oui | non |
| g - une exposition au froid ? | oui | non |
| h - une exposition au vent ? | oui | non |
| i - le matin au lever ? | oui | non |
| j - le soir au coucher ? | oui | non |
| k - la position penchée en avant (par exemple pour mettre ses chaussures) ? | oui | non |
| | | |
| 5 - <u>Est-ce que la douleur vous a déjà réveillé la nuit ?</u> | oui | non |
| | | |
| 6 - <u>Est-ce que la douleur augmente :</u> | | |
| a - en respirant profondément ? | oui | non |
| b - en toussant ? | oui | non |
| c - avec les changements de position du thorax ? | oui | non |
| d - avec les mouvements des épaules et des bras ? | oui | non |
| | | |
| 7 - <u>La douleur entraîne :</u> | | |
| a - une simple gêne ? | oui | non |
| b - une oppression ? | oui | non |
| c - une angoisse ? | oui | non |
| d - une sensation de mort imminente ? | oui | non |
| | | |
| 8 - <u>La douleur est comparable à :</u> | | |
| a - une crampe ? | oui | non |
| b - un coup de poing ? | oui | non |
| c - des piqûres d'aiguilles ? | oui | non |
| d - un étau ? | oui | non |
| e - un corset ? | oui | non |
| f - une brûlure ? | oui | non |
| g - un coup de griffe ? | oui | non |
| h - un pincement ? | oui | non |
| i - une barre ? | oui | non |
| | | |
| 9 - <u>La douleur dure :</u> | | |
| a - une seconde ? | oui | non |
| b - moins d'une minute ? | oui | non |
| c - une à 5 minutes ? | oui | non |
| d - de 5 à 10 minutes ? | oui | non |
| e - de 10 à 15 minutes ? | oui | non |
| f - plus de 15 minutes ? | oui | non |
| g - des heures entières ? | oui | non |
| h - des journées entières ? | oui | non |
| | | |
| 10 - <u>Pour faire céder la crise, avez-vous déjà utilisé un des médicaments suivant :</u> | | |
| a - TRINITRINE ? | oui | non |
| b - NATIROSE ? | oui | non |
| c - RISORDAN ? | oui | non |
| d - ELBETRINE ? | oui | non |
| | | |
| <u>Si oui, avez-vous :</u> | | |
| a - croqué et laissé fondre sous la langue ce médicament ? | oui | non |
| b - avalé sans croquer ce médicament ? | oui | non |

Si oui, la crise a cédé en :

- | | | |
|--------------------------|-----|-----|
| a - 30 secondes ? | | |
| b - 1 à 3 minutes ? | oui | non |
| c - 3 à 10 minutes ? | oui | non |
| d - plus de 10 minutes ? | oui | non |
| e - n'a pas cédé ? | oui | non |

11 - La première fois ou vous avez ressenti cette douleur c'était il y a :

- | | | |
|----------------------------|-----|-----|
| a - plus de 10 ans ? | | |
| b - plus de 5 ans ? | oui | non |
| c - entre 1 et 5 ans ? | oui | non |
| d - entre 6 mois et 1 an ? | oui | non |
| e - entre 3 et 6 mois ? | oui | non |
| f - entre 1 et 3 mois ? | oui | non |
| g - moins d'1 mois ? | oui | non |
| h - moins d'1 semaine ? | oui | non |

12 - Combien de fois souffrez-vous actuellement ?

- | | | |
|-------------------------------------|-----|-----|
| a - plus de 3 fois par jour ? | oui | non |
| b - au moins une fois par jour ? | oui | non |
| c - au moins une fois par semaine ? | oui | non |
| d - au moins une fois par mois ? | oui | non |
| e - moins d'une fois par mois ? | oui | non |
| f - une seule douleur en tout ? | oui | non |

13 - La dernière douleur remonte à :

- | | | |
|---------------------------|-----|-----|
| a - aujourd'hui ? | oui | non |
| b - moins de 3 jours ? | oui | non |
| c - moins d'une semaine ? | oui | non |
| d - moins d'un mois ? | oui | non |
| e - moins de 3 mois ? | oui | non |
| f - moins de 6 mois ? | oui | non |
| g - moins d'un an ? | oui | non |
| h - plus d'un an ? | oui | non |

14 - Est-ce que la douleur a changé depuis un mois ?

Si oui

- | | | |
|---|-----|-----|
| a - les crises sont plus fréquentes ? | oui | non |
| b - les crises surviennent pour des efforts moins importants | oui | non |
| c - les crises surviennent la nuit ? | oui | non |
| d - les crises durent plus longtemps ? | oui | non |
| e - les crises disparaissent moins vite avec le
TRINITRINE ou le RISORDAN, L'ELBETRINE ou la NATRIROSE ? | oui | non |

15 - Combien de fois avez-vous déjà été hospitalisé pour des douleurs dans la poitrine ?

- | | | |
|----------------------|-----|-----|
| a - jamais ? | | |
| b - une fois ? | oui | non |
| c - deux fois ? | oui | non |
| d - 3 à 5 fois ? | oui | non |
| e - plus de 5 fois ? | oui | non |

ANNEXE II

ANNEXE : Modifications des règles

Règles initiales	: Règles construites à l'étape I (GBSV)
Etape II	: Modifications sur la base des résultats de l'évaluation.
Etape III	: Modifications sur la base des résultats des analyses bivariées.
Colonne 1	: Prémises des règles.
Colonne 2	: Effectifs correspondants
Colonne 3	: Valeur du coefficient de vraisemblance à l'étape I
Colonne 4	: Valeur du coefficient de vraisemblance à l'étape II
Colonne 5	: Coefficients de vraisemblance calculés à partir des fréquences à l'étape III
Colonne 6	: Décision des experts.

La présence d'un "S" dans une colonne signifie que la règle a été supprimée.

GROUPE "STAPHYLOCOQUE"

I - Règles initiales	Effectifs	CV1	CV2	CV3	Retenu
BPCO et AB antérieure	15	0,3	idem	S	oui
B Pneumopathie et AB ant.	13	0,3	"	S	0,3
Acquisition hospitalière	155	0,2	S	0,2	oui
Acquisition médecine	102	0,5	idem	S	oui
Acquisition réanimation	28	0,5	0,7	0,1	oui
Diabète	73	0,3	idem	S	0,1
Endocardite	72	0,3	idem	0,1	oui
Infection cutanée	16	0,8	idem	0,5	oui
Infection urinaire	105	0,1	idem	S	oui
Porte entrée catheter	40	0,7	idem	0,5	oui
Porte entrée urinaire	130	0,3	idem	S	oui
Porte entrée cutanée	64	0,7	idem	0,5	oui
Prothèse cardiaque et t° 38,5	16	0,3	idem	0,2	oui
Spondylodiscite	18	0,4	idem	S	oui
Thrombophlébite suppurée	9	0,2	0,8	idem	oui

II - Règles ajoutées à l'étape II

Granulopenie	22		0,5	S	oui
Hémodialyse	1		0,5	idem	oui

III - Règles proposées à l'étape III

Corticoides	22		0,2		oui
Cirrhose	42		0,1		oui
Chimiothérapie	14		0,2		oui
Plaie infectée	30		0,1		oui
Porte entree colique	48		-0,9		oui
Sonde urinaire	17		0,1		oui
Purpura	20		0,2		oui
Valvulopathie et t° 37°5	49		0,2		oui
Valvulopathie et t° 37°5	13		0,1		oui
1 âge 15	22		0,1		oui

GROUPE "PENI-S"

I - Règles initiales	Effectif	CV1	CV2	CV3	retenu
Asplénie et nourrisson	0	0,3	idem	S	oui
BPCO et non AB-antérieure	30	0,4	idem	0,3	oui
Bpneumopathie	39	0,5	idem	0,3	oui
Bpneumopathie et BPCO	14	0,5	idem	S	oui
Endocardite	72	0,6	idem	0,3	oui
Endocardite et inf. urinaire	0	0,6	idem	S	oui
Infection cutanée	20	0,5	idem	0,1	oui
Infection urinaire	105	0,4	idem	S	oui
Méningite et âge 40	37	0,7	idem	0,4	oui
Non acquisition hospitalière	400	0,2	idem	0,1	oui
Pneumonie alvéolaire	58	0,8	idem	0,7	oui
Porte entrée colique	48	0,7	idem	S	oui
Porte entrée cutanée	64	0,6	idem	S	0,2
Prothèse cardiaque et t° 38	20	0,2	idem	S	oui
Purpura	19	0,2	idem	0,4	oui
Sonde urinaire	17	0,2	idem	S	oui

II - Règles ajoutées à l'étape II

AB antérieure	223	-0,3	S	oui
Sida	9	0,2	S	oui

III - Règles proposées à l'étape III

Insuffisance cardiaque	52		0,1	non(S)
Pathologie colorectale	20		0,2	oui
Porte entrée catheter	40		0,3	oui
Asplénie et âge 15	1		0,3	oui

GROUPE "AMPI-S"

I - Règles initiales	Effectifs	CV1	CV2	CV3	retenu
Hémopathie	26	0,3	idem	S	oui
Non acquisition hospitalière	400	0,1	idem	S	oui
Porte entree urinaire	130	0,3	idem	S	0,1

II - Règles ajoutées à l'étape II

Infection urinaire	105	0,5	S	oui
Paraplégie	1	0,3	idem	oui
Sonde urinaire	17	0,5	S	0,1

III - Règles proposées à l'étape III

Endocardite	72		.2	oui
Pathologie colorectale	20		.1	oui
Pneumonie alveolaire	58		-0,9	oui
Porte entrée catheter	40		.2	oui
Sida	9		.2	oui

GROUPE "BGN Ville"

I - Règles initiales	Effectifs	CV1	CV2	CV3	retenu
Arthrite et acquisition hospitalière	5	0,3	idem	S	oui
Asplénie et nourisson	0	0,3	idem	S	oui
BPneumopathie et BPCO	14	0,5	idem	S	0,1
Acquisition chirurgie	53	0,4	0,2	S	oui
Acquisition chirurgie digestive	9	0,8	0,2	idem	oui
Acquisition hématologie	5	0,3	5	idem	oui
Acquisition médecine	102	0,5	idem	S	oui
Acquisition réanimation	28	0,5	-0,5	S	oui
Acquisition urologie	3	0,8	0,4	S	oui
Diabète	73	0,4	idem	S	0,2
Granulopénie	22	0,5	-0,5		
Immunodépression	29	0,5	5		
LLC	5	0,4	5		
Lymphome	8	0,5	5		
Méningite et 0,1 âge 15	2	0,8	5		
Neutropénie	21	0,8	0,5	S	oui
Non acquisition hospitalière	400	0,25	5		
Otélite	17	0,2	idem	S	oui
Ostéite et prothèse os	1	0,3	0,1	idem	oui
Otite	3	0,5	5		
Péritonite	1	0,8	0,6		
Pneumonie alvéolaire	58	0,5	0,2	0,1	oui
Pneumonie alveolaire et BPCO	6	0,5	0,3	S	oui
Porte entrée catheter	40	0,2	idem	-0,4	oui
Porte entrée colique	48	0,7	idem	S	oui
Sonde urinaire	17	0,5	idem	0,3	oui
Spondylodiscite	18	0,4	idem	S	oui
Tumeur intestinale	10	0,5	idem	0,2	oui
Tumeur solide et chimio	7	0,5	5		

II - Règles ajoutées à l'étape II

Infection cutanée	16		0,1	S	oui
Sida	9		0,4	0,2	oui
AB antérieure	223		-0,2	S	oui

III - Règles proposées à l'étape III

Age 65	310			0,1	oui
Endocardite	72			-0,9	oui
Leucopénie	42			0,1	oui
Lithiase biliaire	20			0,5	oui
Méningite et âge 40	37			-0,5	oui
Risque professionnel = 4	45			0,3	oui
Uropathie	31			0,6	oui
Bpneumopathie	39			-0,5	oui

GROUPE "SALMONELLE"

I - Règles initiales	Effectifs	CV1	CV2	CV3	retenu
Conscience anormale	143	0,2	idem	S	oui
Leucopenie et VS 50	14	0,5	idem	0,3	oui
LLC	5	0,2	0,6	idem	oui
Ostéite	17	0,1	idem	S	oui
Ostéite et drépanocytose	0	0,8	idem	S	oui
Ostéomyélite et drépanocytose	0	0,9	idem	S	oui
Porte entrée colique	48	0,5	idem	0,3	oui
Sida	9	0,5	0,9	0,2	oui
Syndrome dysentérique	14	0,5	idem	0,9	oui

II - Règles ajoutées à l'étape II

Granulopénie	22	0,4	S	oui
Lymphopénie	15	0,4	S	oui
Pneumonie alvéolaire	58	-0,9	idem	oui

III - Règles proposées à l'étape III

Age 14	15	0,2	oui
Température 38,5	31	-0,4	oui
Diabète	72	-0,9	oui
Endocardite	29	-0,1	oui
Immunodépression	130	-0,9	oui
Porte entrée urinaire	0	-0,9	oui
Infection os et drépanocytose			

GROUPE "BGN HOSPITALIER"

I - Règles initiales	Effectifs	CV1	CV2	CV3	retenu
Acquisition chirurgie	53	0,4	idem	S	oui
Acquisition hematologie	5	0,2	0,3	idem	oui
Acquisition médecine	102	0,3	idem	0,1	oui
Acquisition réanimation	28	0,5	0,7	0,2	oui
Arthrite et acquisition hospitalière	5	0,5	idem	S	oui
BPCO et AB antérieure	15	0,3	5	0,1	oui
BPneumopathie et AB antérieure	13	0,3	idem	0,2	oui
BPneumopathie	39	0,3	idem	0,1	oui
Choc Toxi-infectieux	2	0,7	0,8	idem	oui
Bpneumopathie et BPCO	14	0,3	idem	S	oui
Diabète	73	0,2	5		oui
Granulopénie	22	0,6	0,3	0,2	oui
Infection cutanée	20	0,3	idem	S	oui
Infection urinaire	105	0,6	idem	S	oui
Non acquisition hospitalière	400	0,1	idem	S	oui
Ostéite	17	0,2	idem	S	oui
Ostéite et prothèse os	1	0,2	0,4	idem	oui
Porte entrée catheter	40	0,2	idem	S	oui
Porte entrée urinaire	130	0,4	0,3	S	oui
Sonde urinaire	17	0,5	idem	0,1	oui

II - Règles ajoutées à l'étape II

Hémodialyse	1	0,1	idem	oui
Lésion colorectale	5	0,1	idem	oui
Meningite	55	0,1	S	oui

III - Règles proposées à l'étape III

Ascite	26	0,2	oui
Chimiothérapie	14	0,3	oui
Acquisition hospitalière	155	0,2	oui
Endocardite	72	-0,9	oui
Hémopathie	26	0,1	oui
Insuffisance rénale	26	0,2	oui
Tumeur intestinale	15	0,5	non

GRUPE "BRUCELLE-YERSINIA"

I - Règles initiales	Effectifs	CV1	CV2	CV3	retenu
Leucopenie et VS 50	14	0,5	idem	S	0,1

II - Règles ajoutées à l'étape II

Acquisition hospitalière	155	-0,8	idem	oui
Endocardite	72	-0,5	idem	oui
GB 13	283	-0,6	idem	oui

Groupe "Anaérobic"

I - Règles initiales

Bpneumopathie	39	0,2	idem	0,1	oui
Porte entrée colique	42	0,5	idem	0,1	oui
Tumeur intestinale	15	0,5	idem	0,1	oui
Tumeur solide et chimio	7	0,3	idem	S	0,1

II - Règles proposées à l'étape III

Porte entrée urinaire	130	-0,9	oui
Radiothérapie	10	0,1	oui

Groupe "Candida"

I - Règles ajoutées à l'étape II Effectifs CV1 CV2 CV3 retenu

Acquisition réanimation	28	0,5	0,1	oui
-------------------------	----	-----	-----	-----

II - Règles proposées à l'étape III

BPCO et AB antérieure	15	0,1	oui
Acquisition médecine	102	0,1	oui
Porte entrée urinaire	130	-0,9	oui
Tumeur solide et chimio	7	0,1	oui
Asthme	15	0,1	non (S)
Pathologie pancréatique	10	0,1	non (S)
Radiothérapie	10	0,2	oui

Groupe "autres germes"

I - Règles initiales Effectifs CV1 CV2 CV3 retenu

BPneumopathie	39	0,2	idem	0,1	oui
Cirrhose	42	0,2	idem	0,1	oui
Immuno-dépression	29	0,3	idem	0,1	oui
Méningite	55	0,2	idem	S	oui
Pneumonie alvéolaire	58	0,2	idem	S	oui

II - Règles ajoutées à l'étape II

Sida	9	0,2	0,2	oui
------	---	-----	-----	-----

III - Règles proposées à l'étape III

Plaie infectée	30	0,1	oui
Porte entrée urinaire	130	-0,9	oui

ANNEXE III

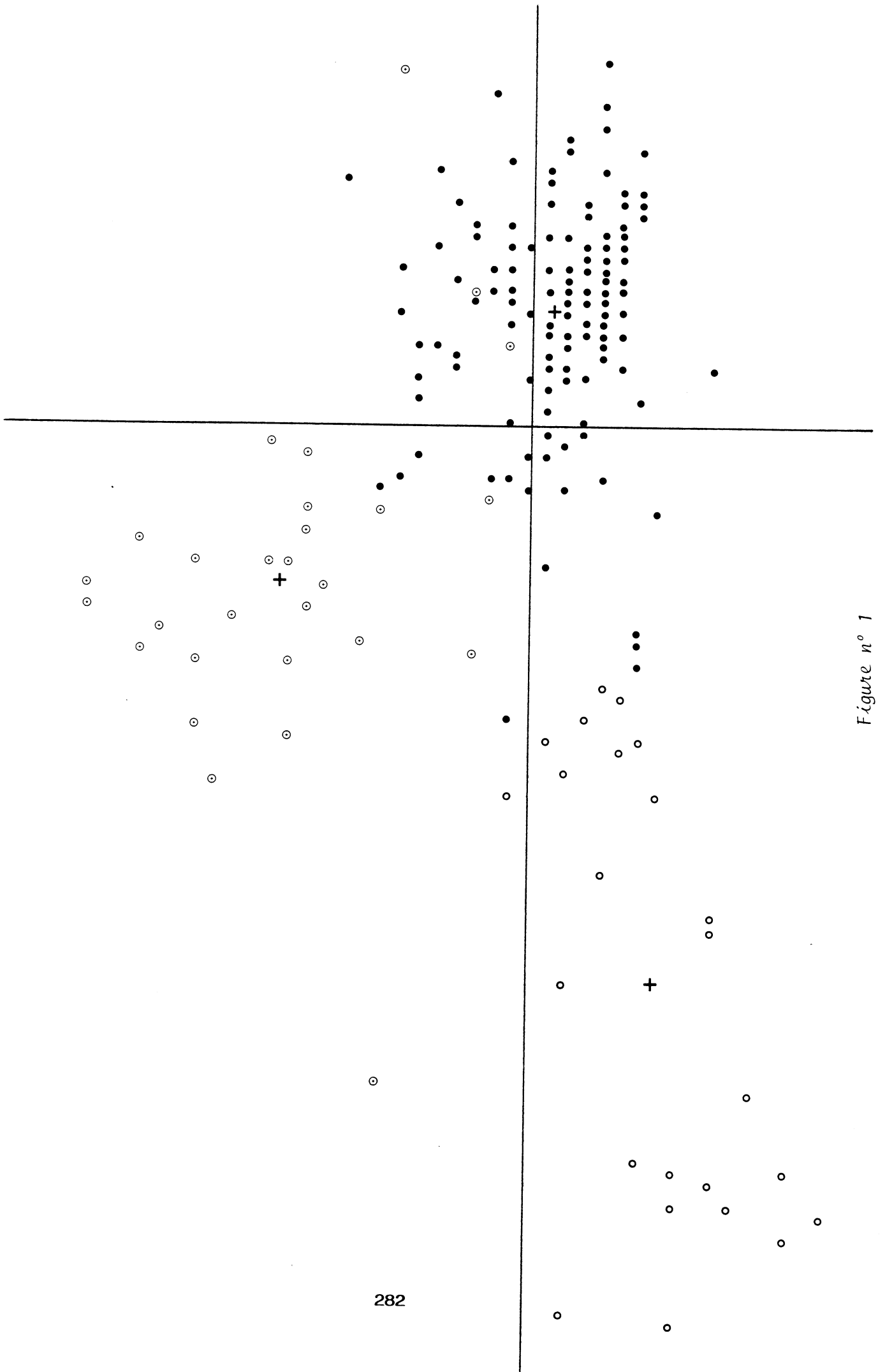


Figure n° 1

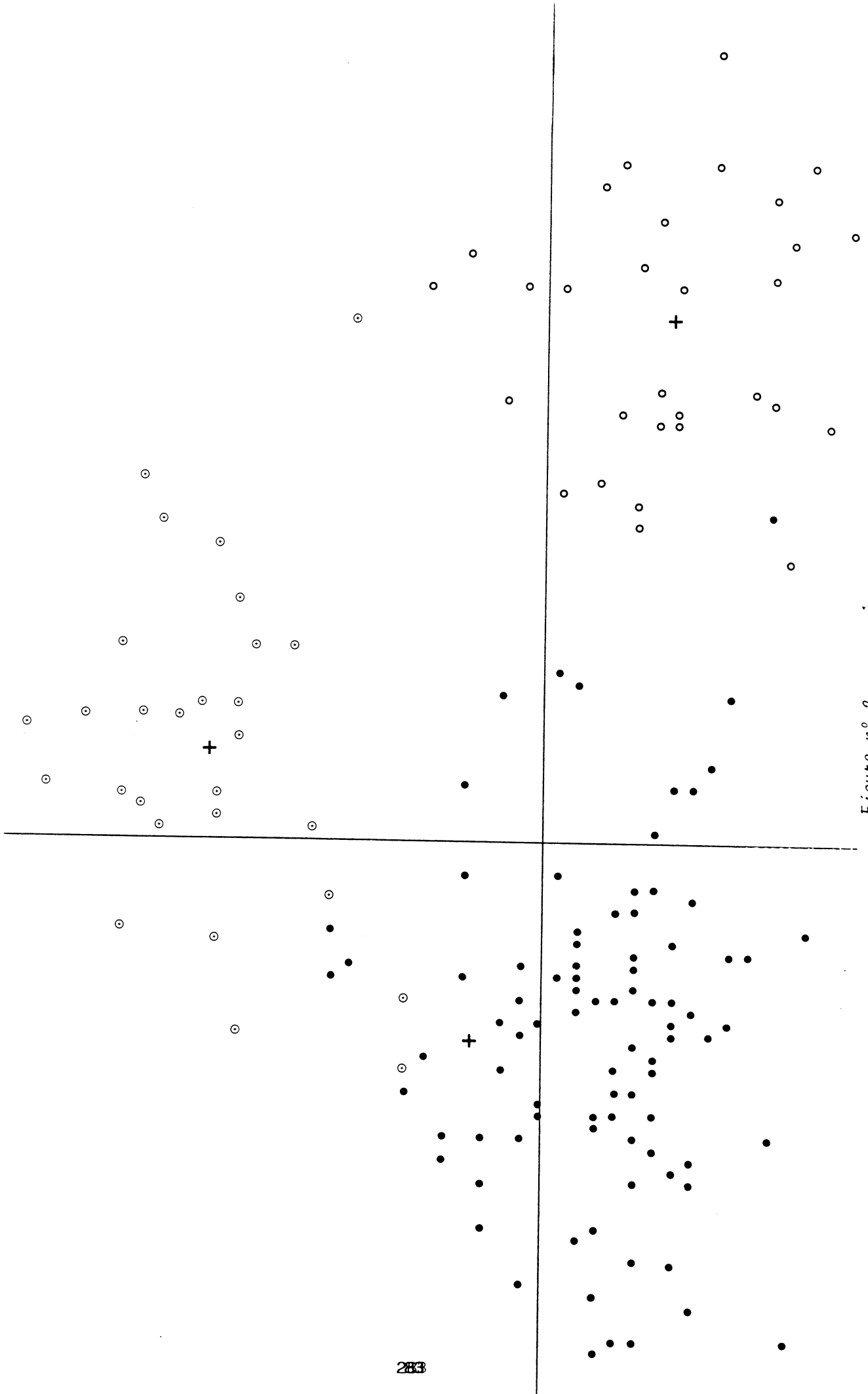


Figure n° 2

• : STAPH

○ : AMPI-S

⊙ : BGN-HOSPITALIER

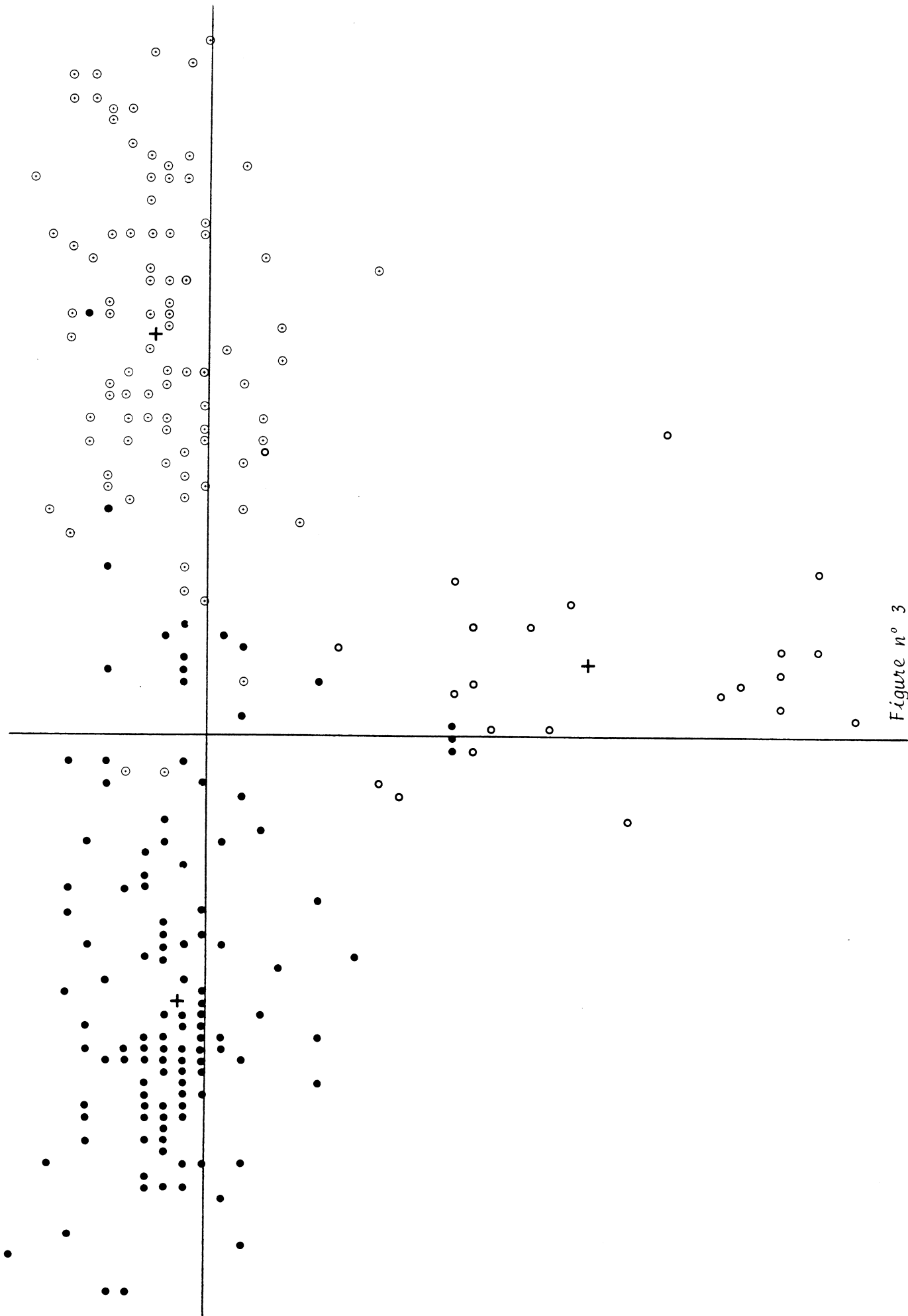


Figure n° 3

● STAPH

○ BCGMUTIF

○ CAMBIEIT

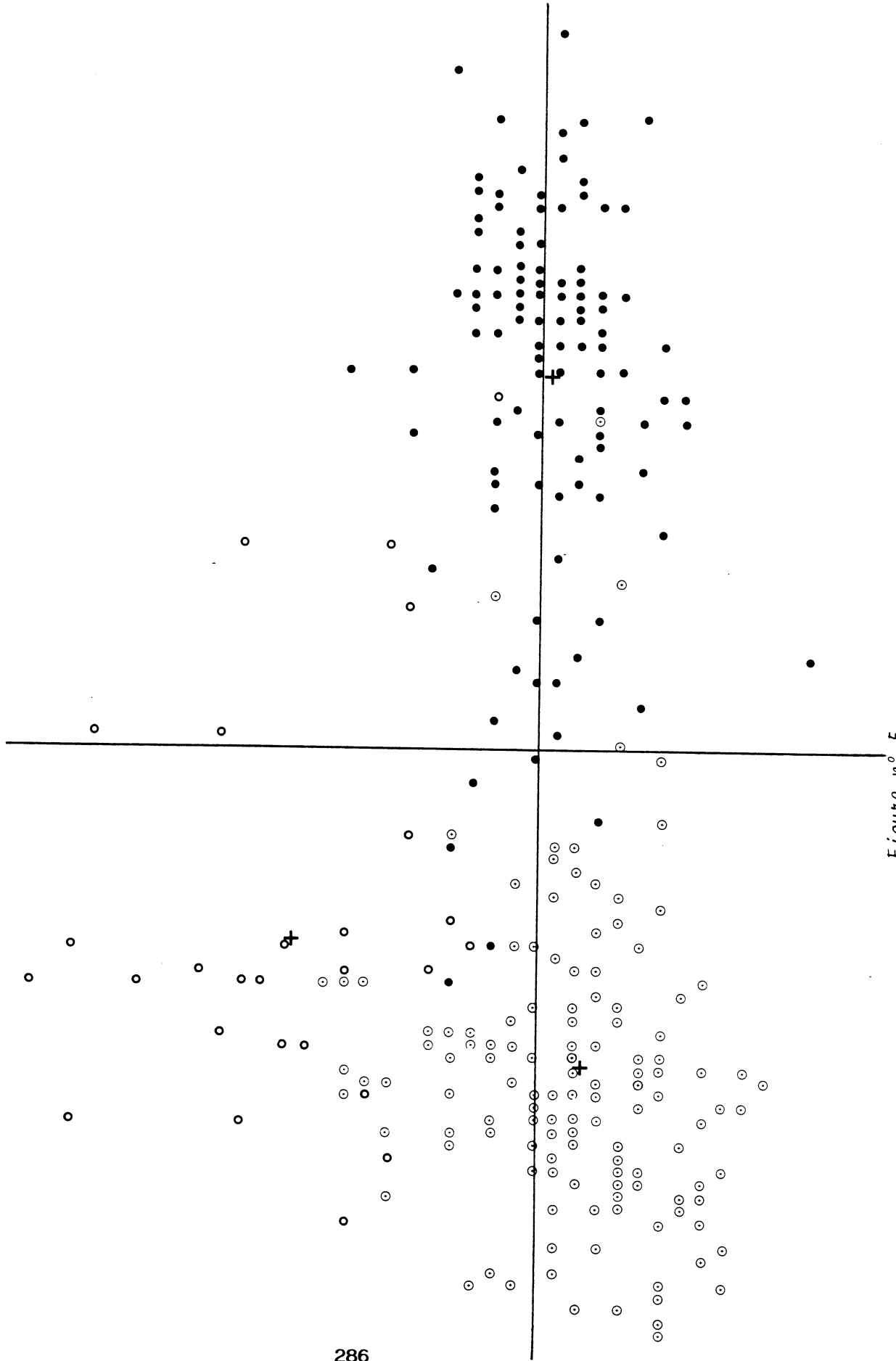


Figure n° 5

⊙ : PENI-S ○ : AMPI-S ● : BGN-VILLE

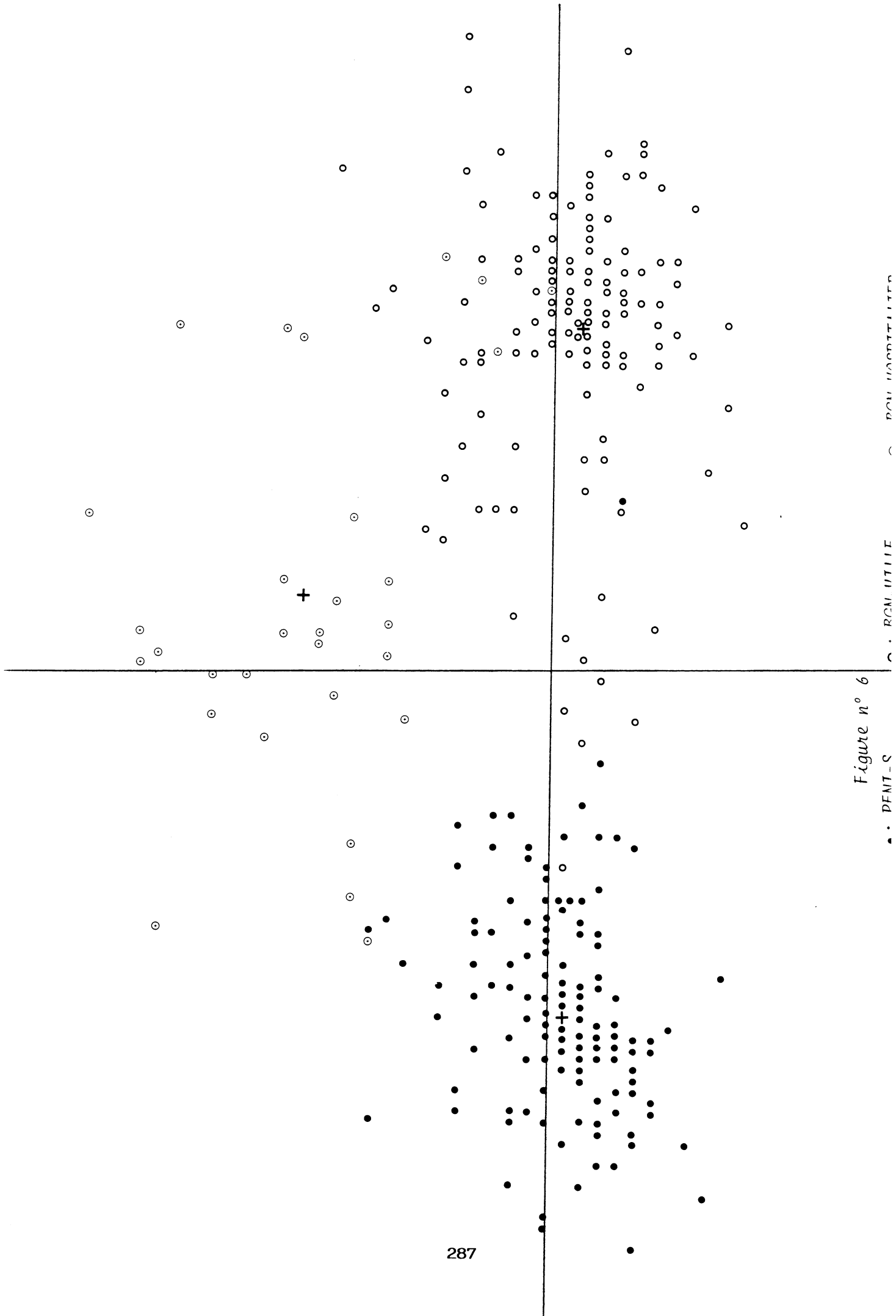


Figure n° 6

• • DFMT-S

○ • RCM UTILITE

○ • RCM UTILITE

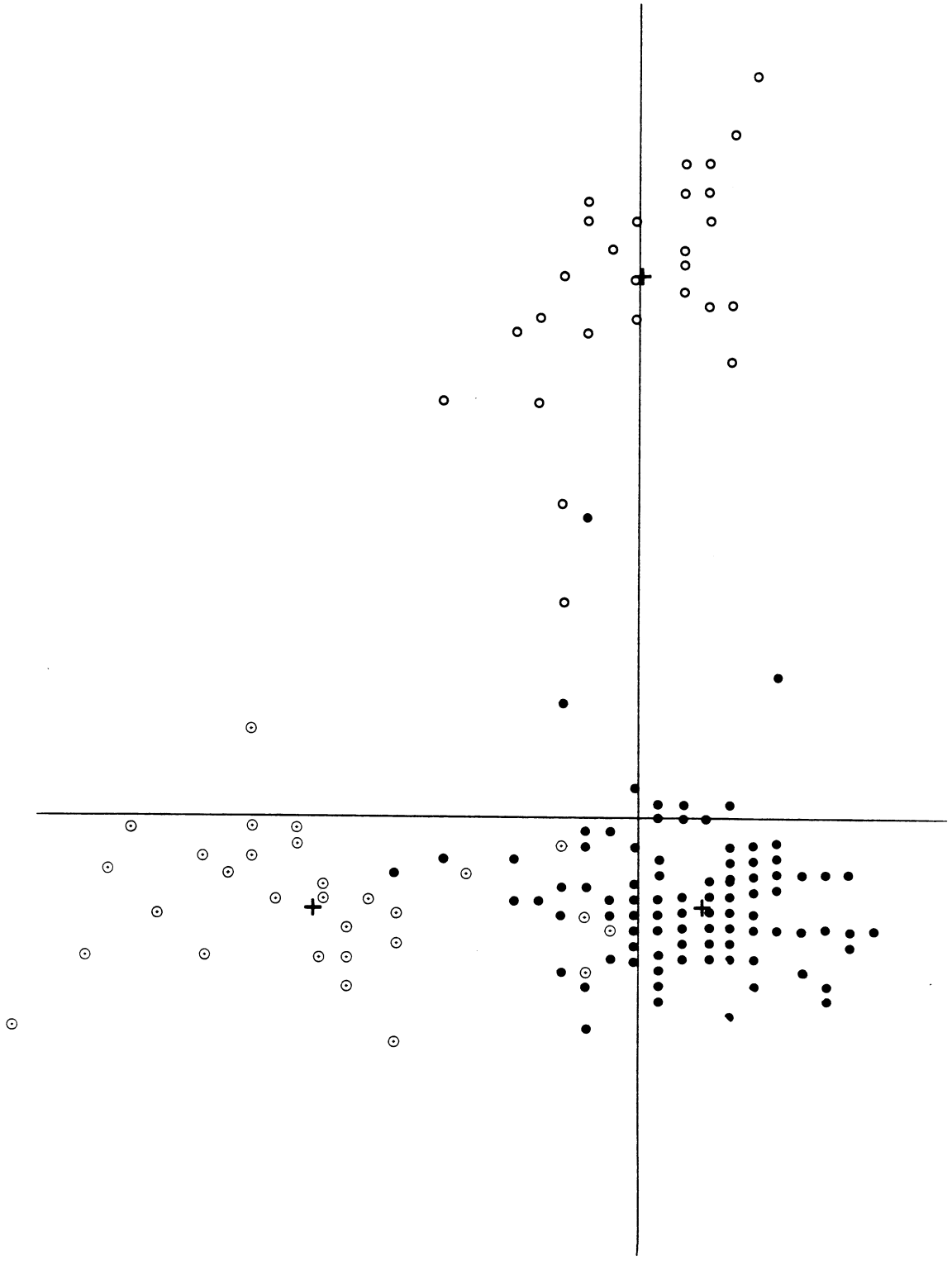


Figure n° 7

○ : AMPI-S

● : BGN-VILLE

⊙ : BGN-HOSPITALIER

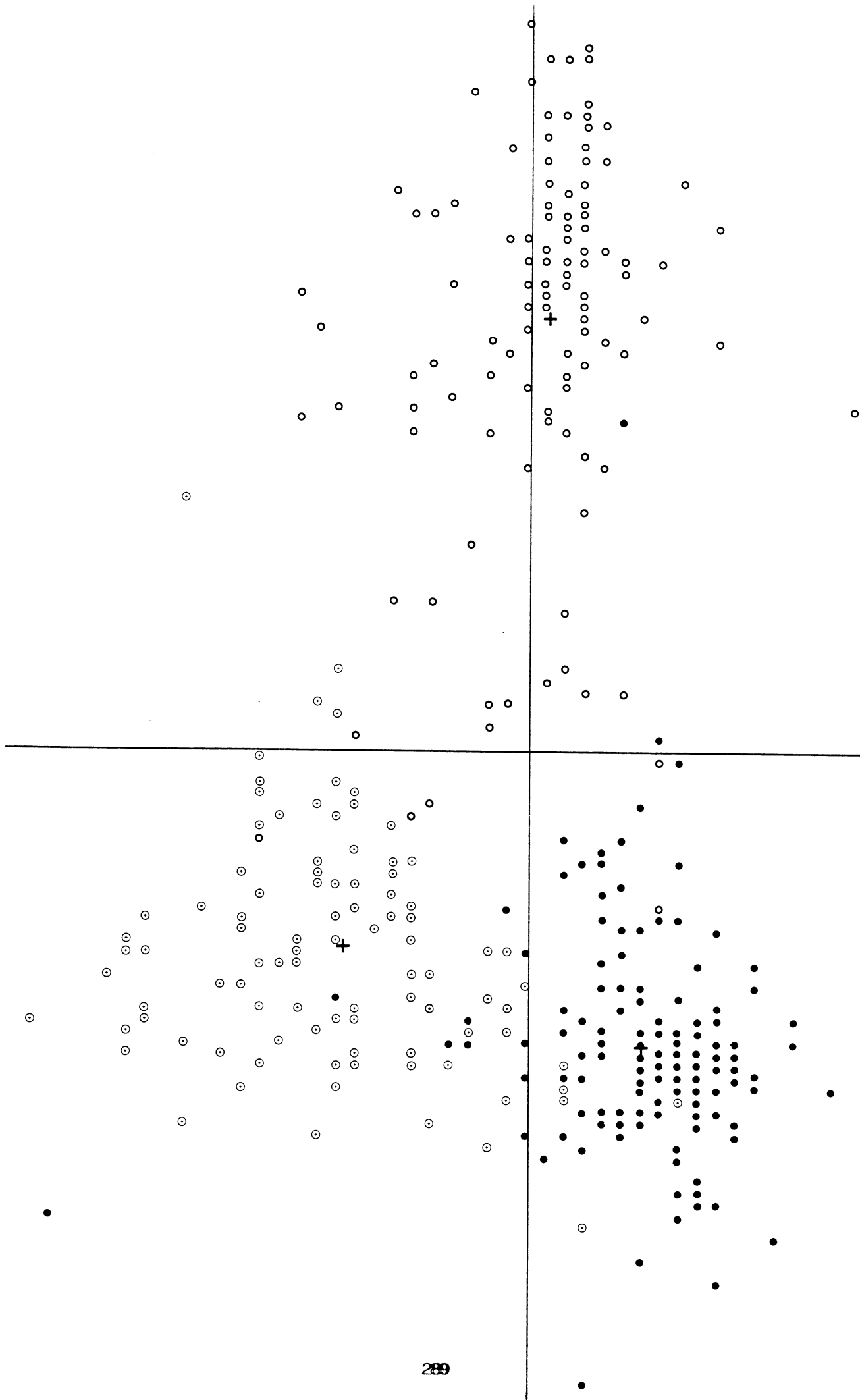


Figure n° 8

○ : BGN-VILLE ⊙ : STAPH ● : PENI-S

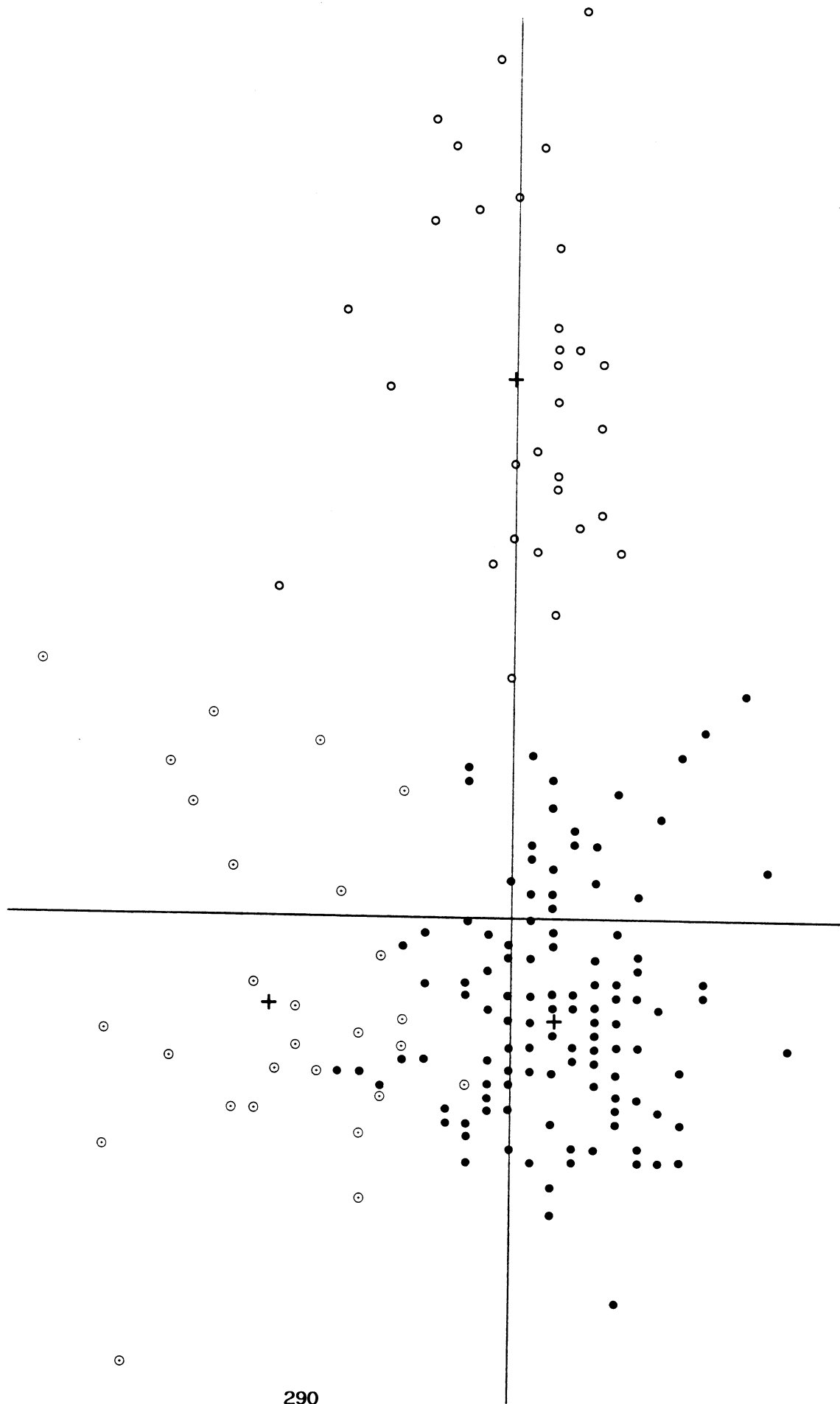


Figure n° 9

• : PENI-S ⊙ : AMPI-S ○ : BGN-HOSPITALIER

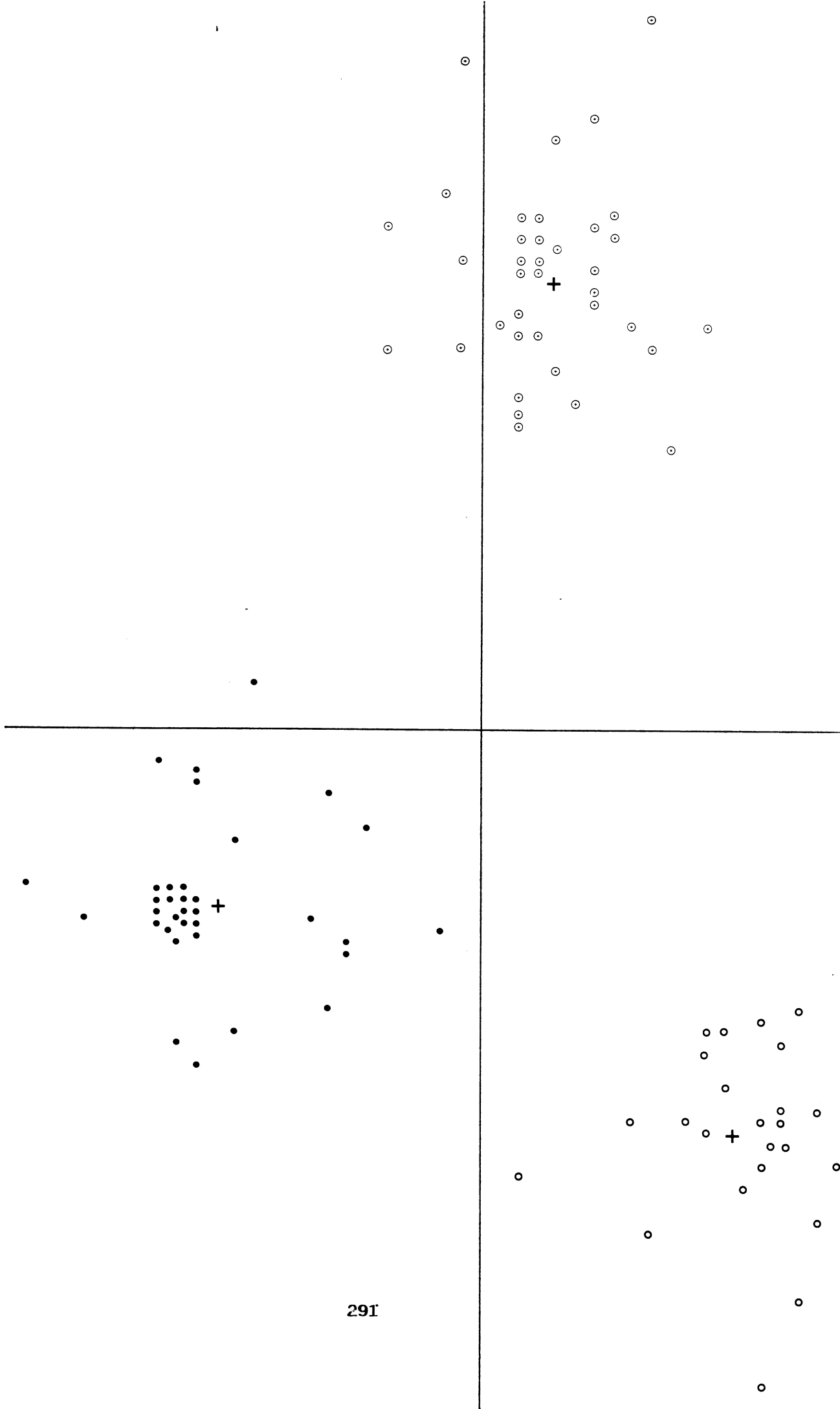


Figure n° 10

• : BGN-Hospit ⊙ : AMPI-S ○ : SALMONELLE

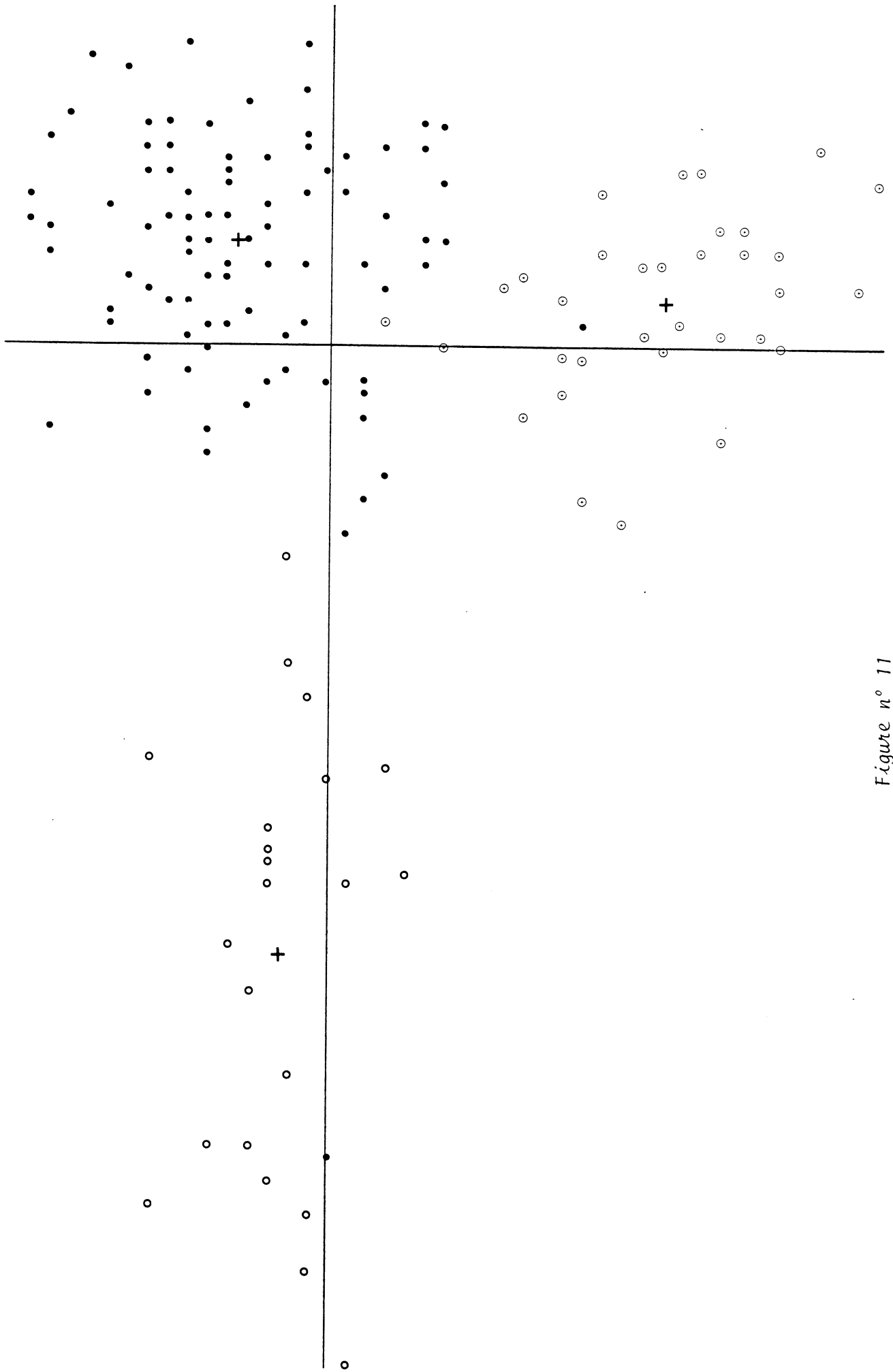


Figure n° 11

• : STAPH ⊙ : AMPI-S ○ : SALMONELLE

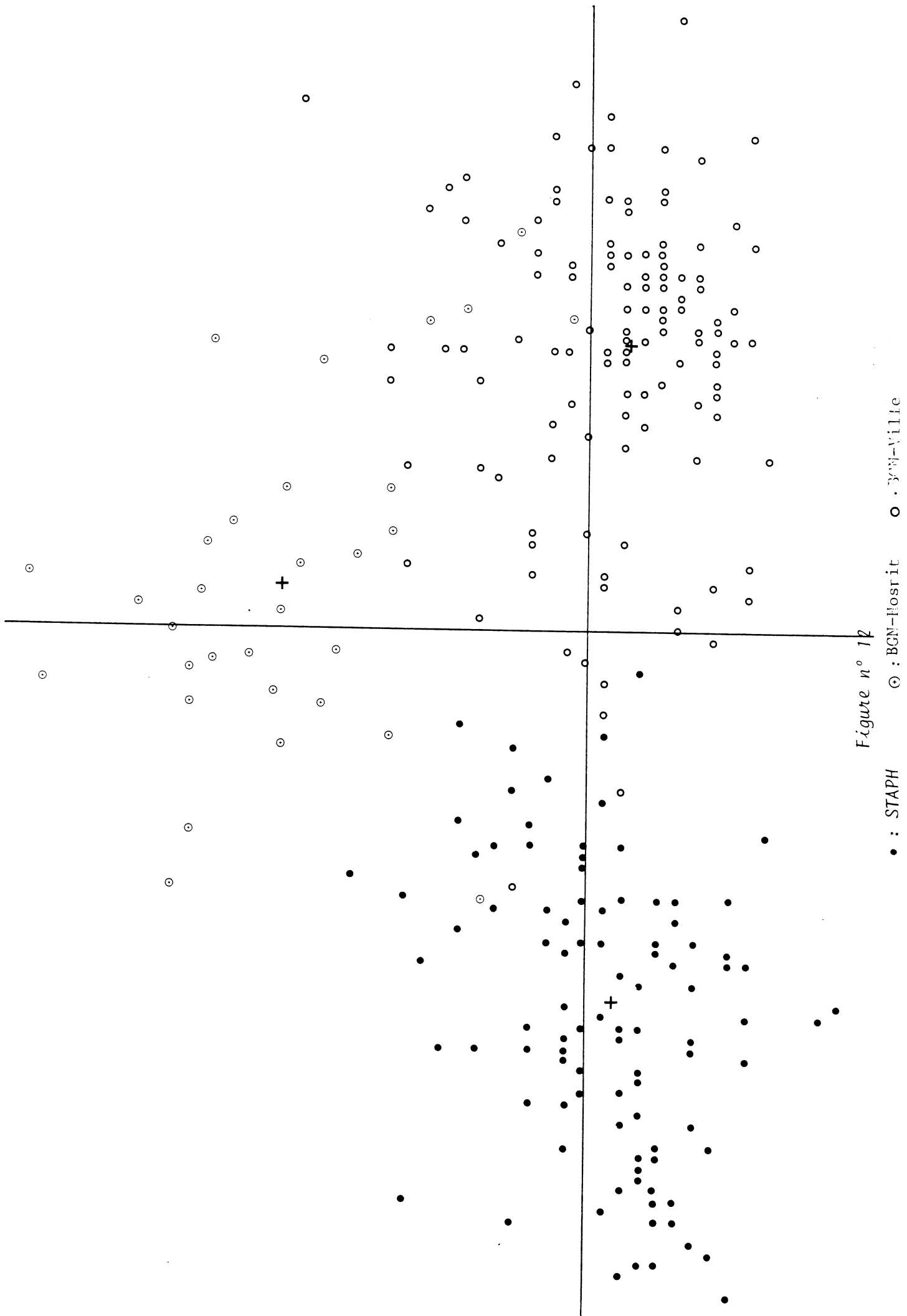


Figure n° 12

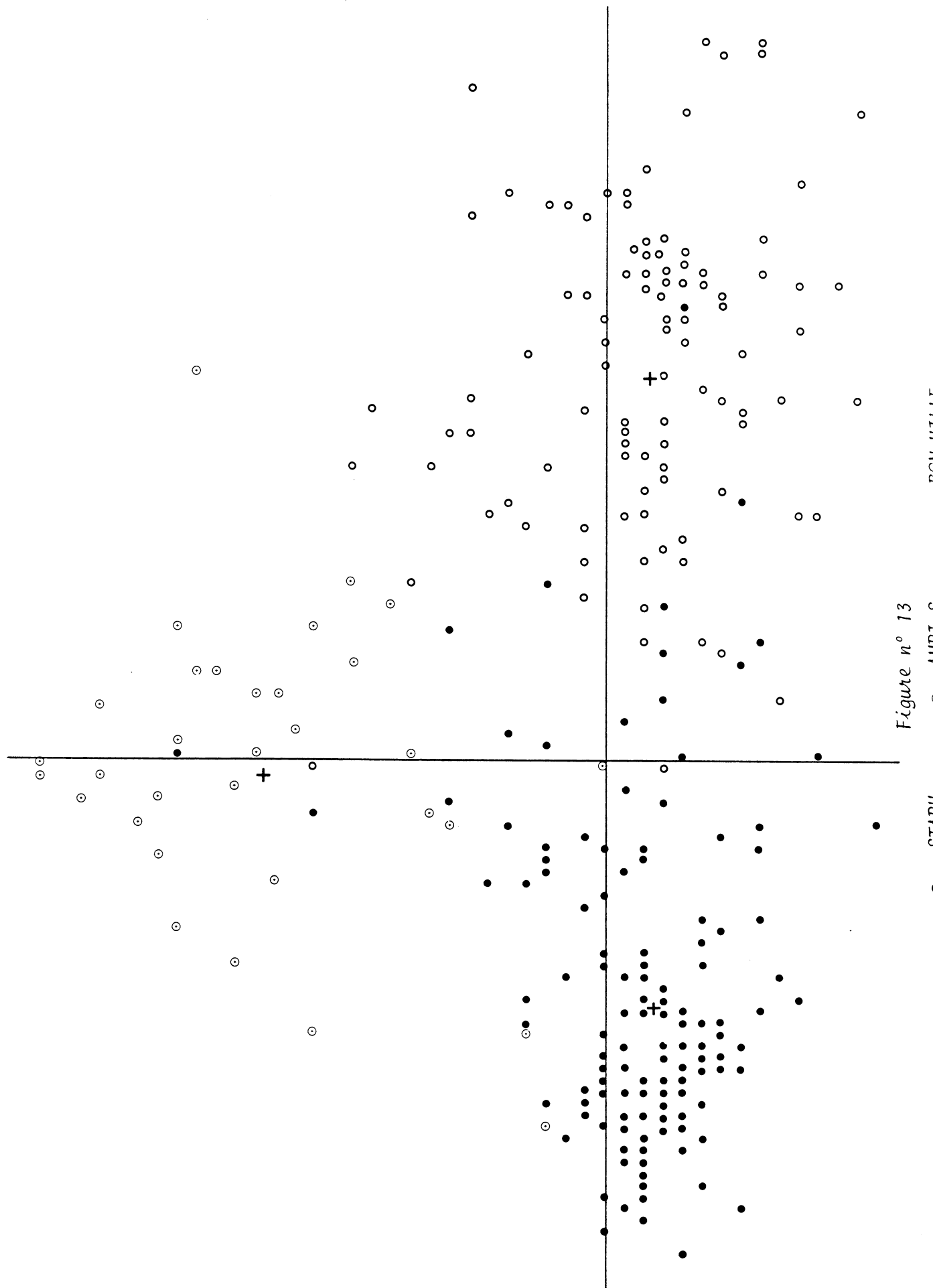


Figure n° 13

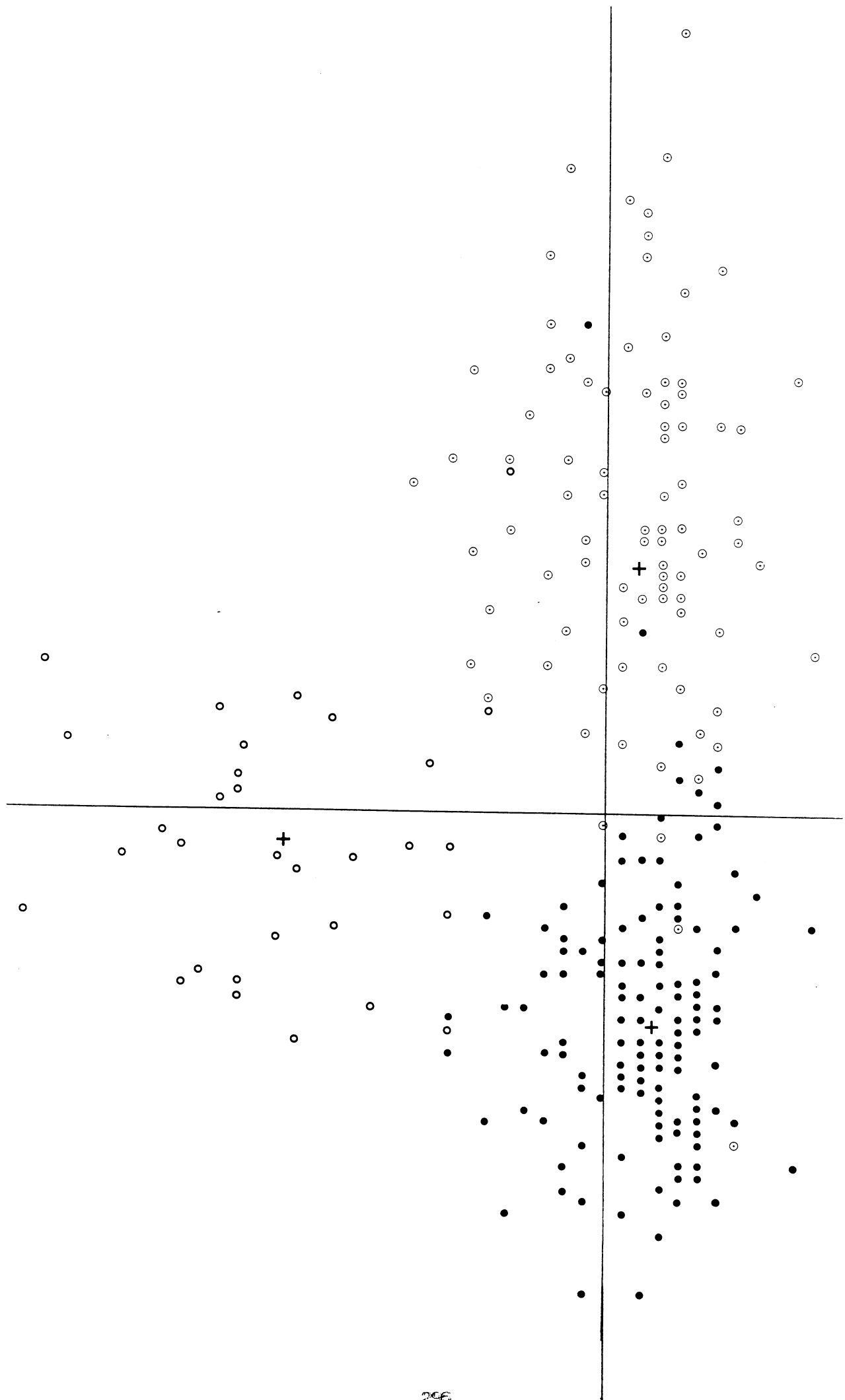


Figure n° 15

○ • STAPH ● • PFMT-S ○ • RCM-HOSPITALISED

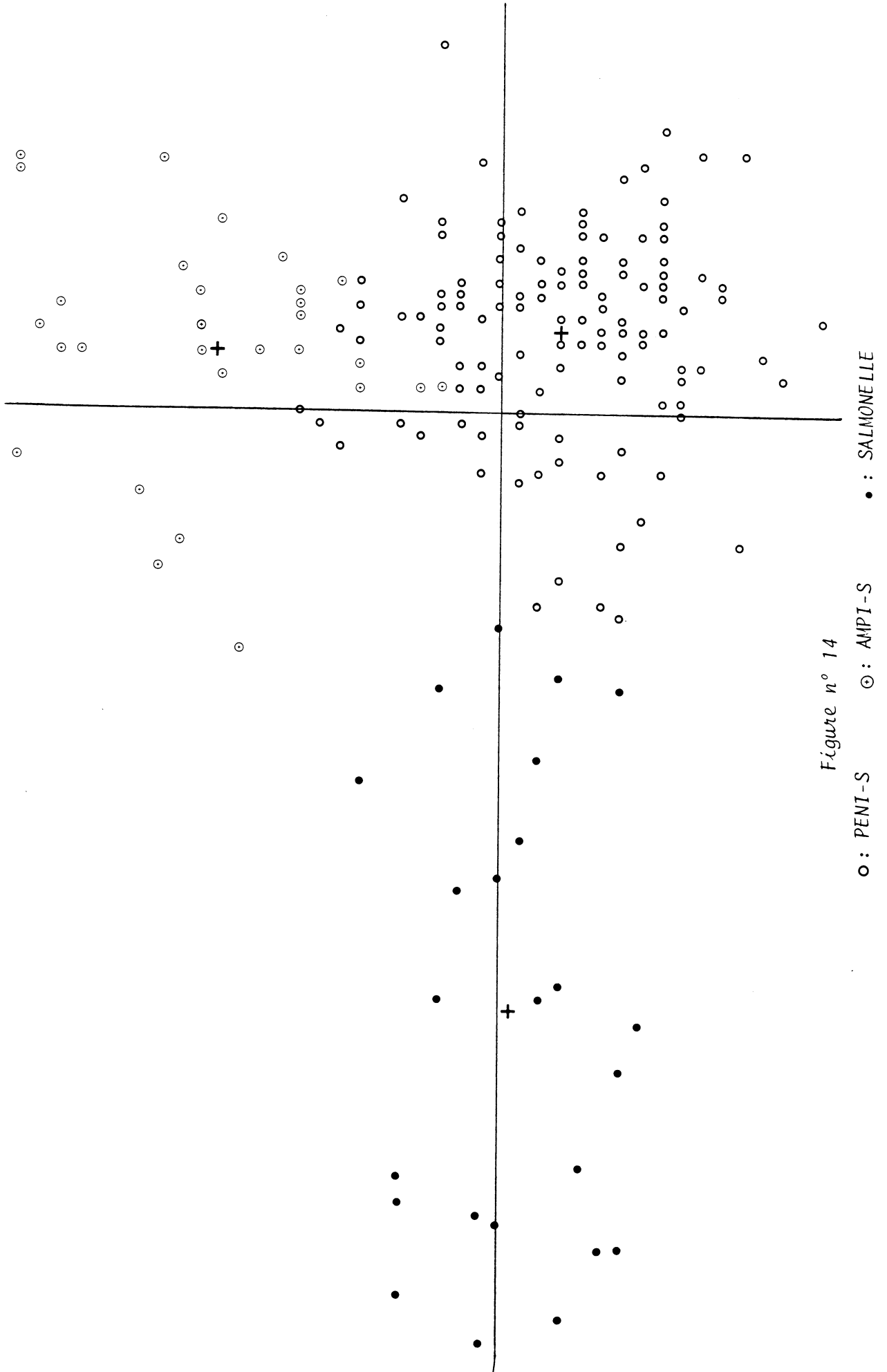


Figure n° 14

○ : PENI-S ⊙ : AMPI-S ● : SALMONELLE

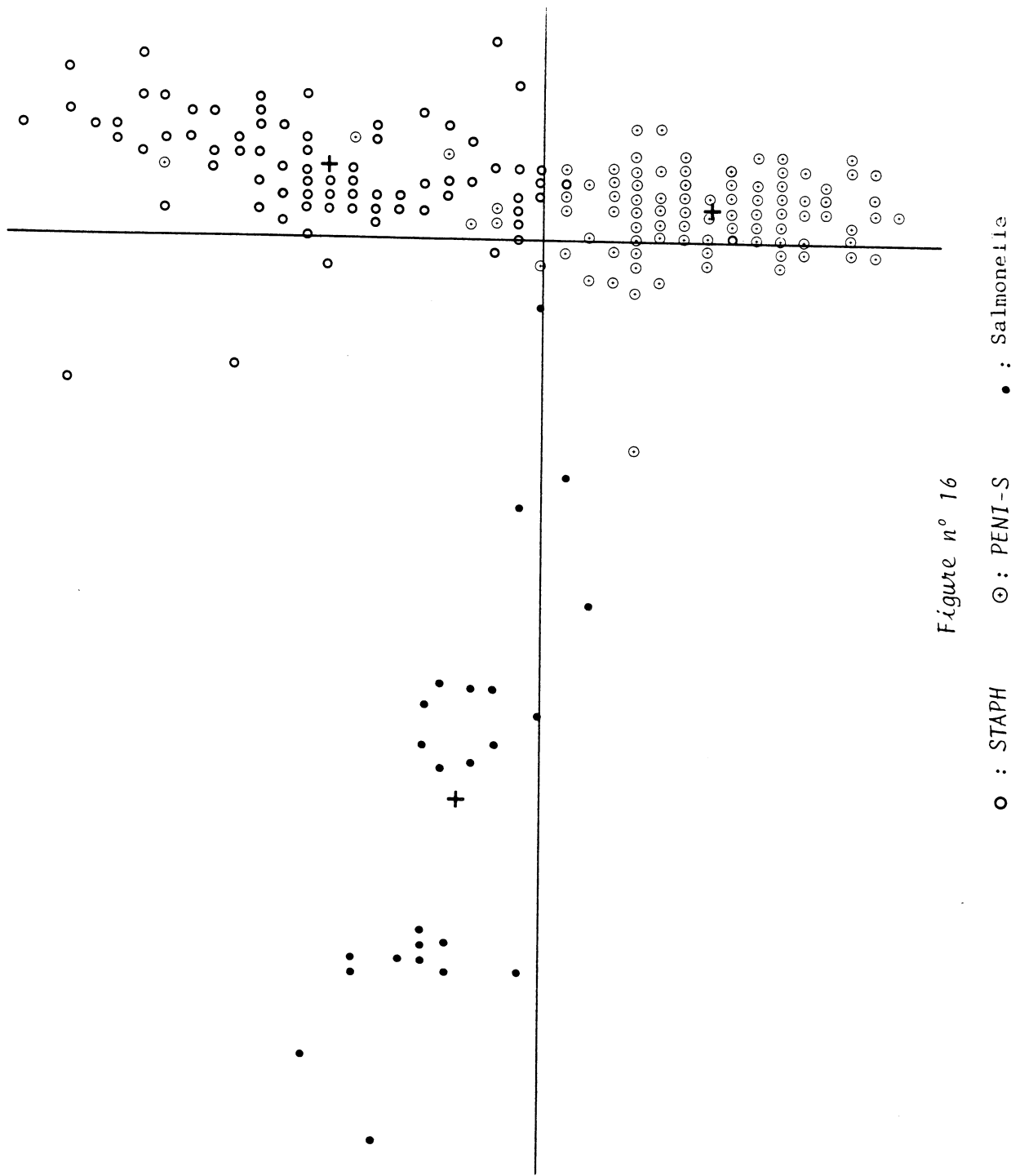


Figure n° 16

O : STAPH ⊙ : PENI-S ● : Salmoneille

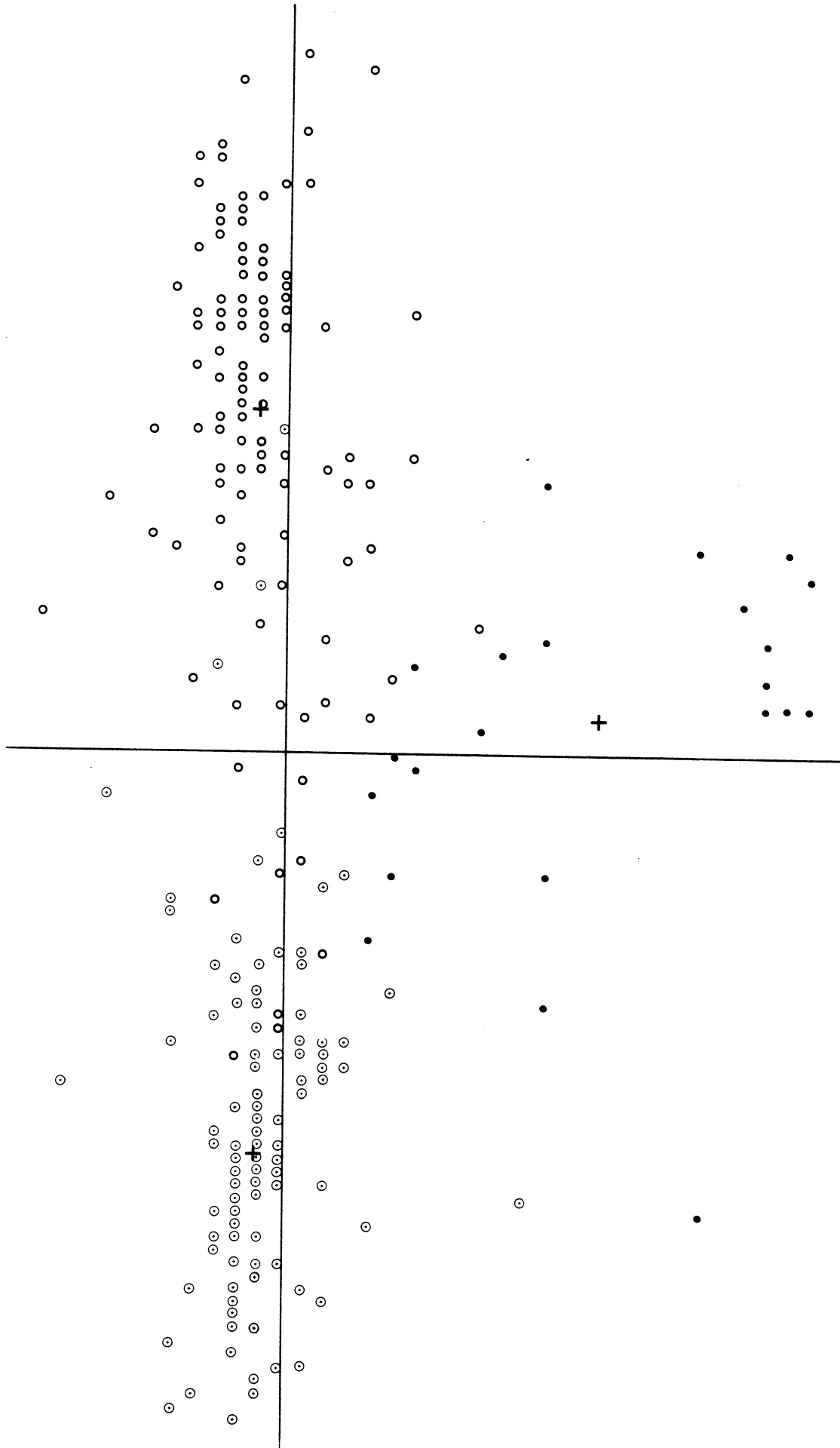


Figure n° 17

⊙ : PENI-S

○ : BGN-VILLE

• : SALMONELLE

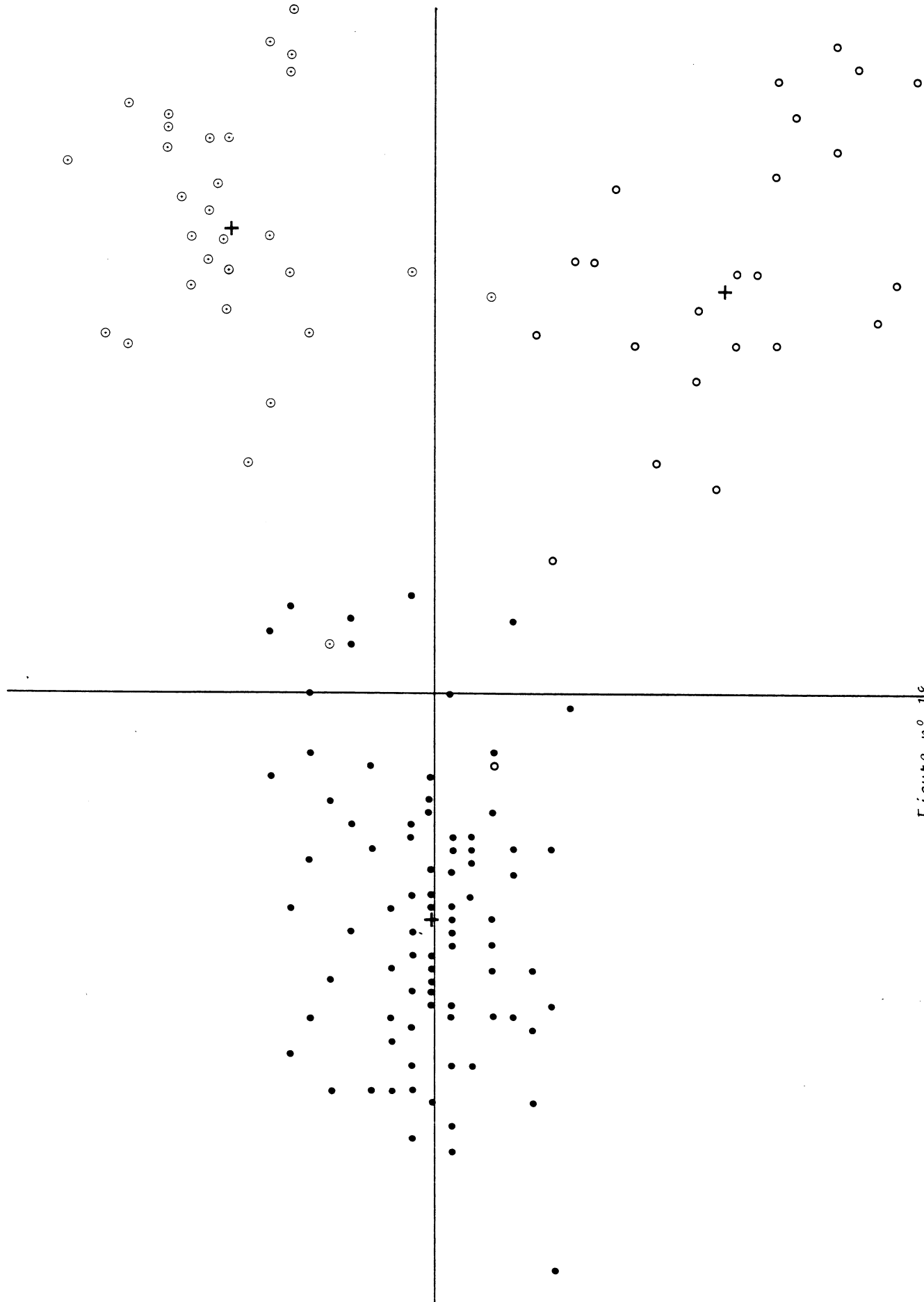


Figure n° 118

● : STAPH ○ : SALMONELLE ⊙ : BGN-HOSPITALIER

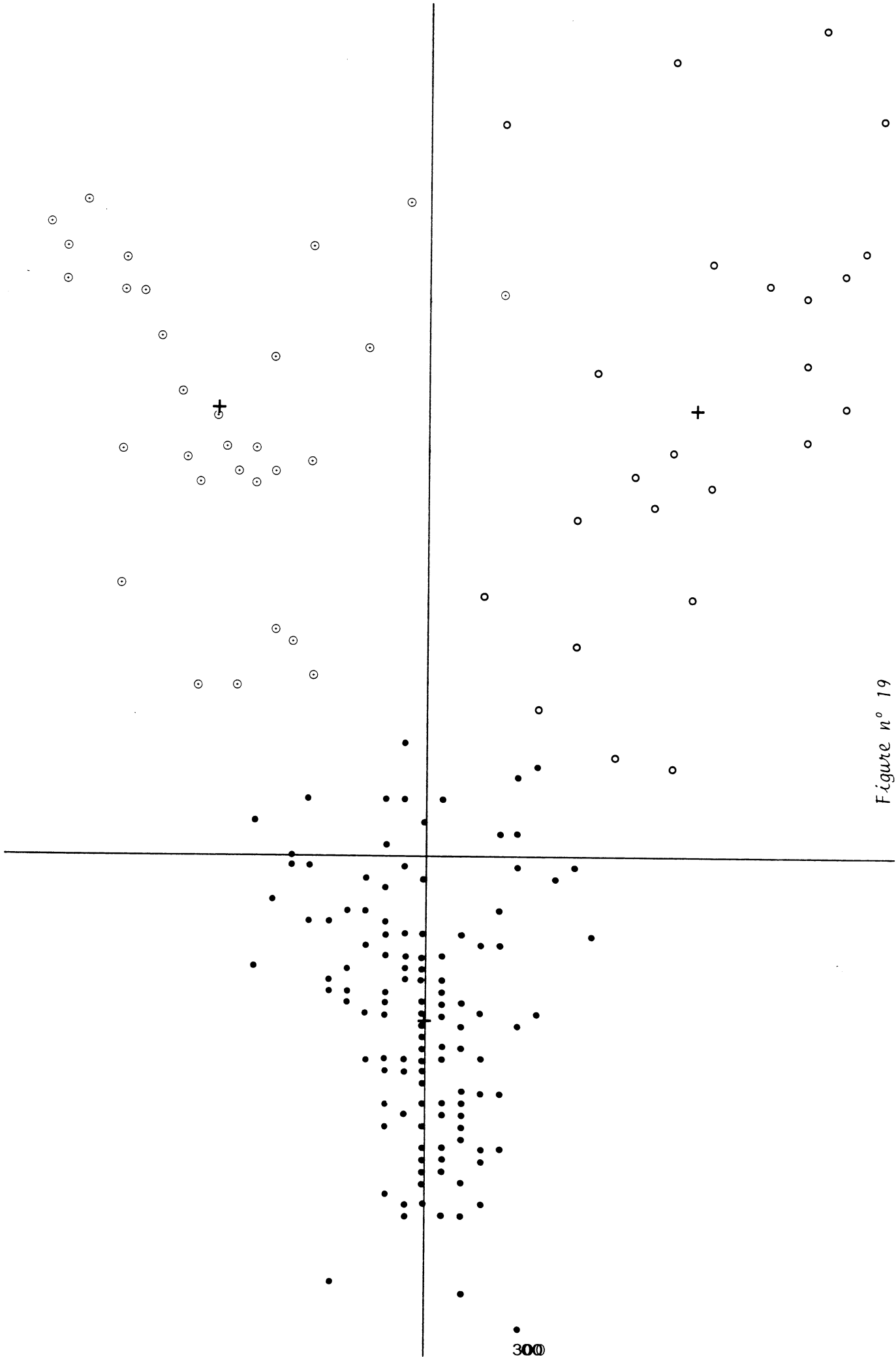


Figure n° 19

BIBLIOGRAPHIE

- (ADLASSNIG 82) ADLASSNING K.P.
A survey of medical diagnosis and fuzzy subsets.
Approximate Reasoning in Decision Analysis, Gupta and Sanches (eds),
North Holland, pp 203-217.
- (BABEAU 85) BABEAU P. and coll
Aspects épidémiologiques actuels des Septicémies.
In : Sem. Hôp. Paris, 61, n°17, pp 1169-1175.
- (BARR 85) BARR A., FEIGENBAUM E.A.
Le livre de l'Intelligence Artificielle.
Eds Eyrolles, Paris 1985.
- (BENZECRI 81) BENZECRI JP and coll
Pratique de l'Analyse des Données.
Eds Dunod, Paris 1981.
- (C. BEUSCART 88) BEUSCART C.
Mise au point d'un Système Expert d'Aide au diagnostic et à la
therapeutique des Septicémies.
Mémoire DEA - GBM - Grenoble, 1988.
- (BEUSCART 89) BEUSCART R.
Aide à l'interprétation des scintigraphies osseuses.
a paraître.
- (BEUSCART 87a) BEUSCART R., BEUSCART C., AUVERGNAT JC, DUHAMEL A.,
MORICE V. et le Groupe SES.
Système Expert Septicémies : expertise multiple et aide à la décision
medicale.
Proc. "IA/SANTE 87" p 115-126.
- (BEUSCART 87b) BEUSCART R., BEUSCART C., DUHAMEL A. et le groupe SES
Septicemia expert system.
Proc. of IEEE/Engineering in Medicine and Biology Society, Nov. 13-16,
1987, BOSTON, pp 929-930.
- (BEUSCART 88a) BEUSCART R., BEUSCART C., DUHAMEL A.
The Septicemia-Expert-System (SES).
Artificial Intelligence and Cognitive Sciences, Manchester University Press,
1988, p 253-262.

- (BEUSCART 88b) BEUSCART R., DUHAMEL A., BEUSCART C., et le groupe SES
Un système d'aide au diagnostic et à la thérapeutique en infectiologie
accessible par réseau télématique.
Proc. EC2. Systèmes Experts et Télématique, Paris, 28-29 Janv. 1988. pp
137-146.
- (BEUSCART 89) BEUSCART R., DUHAMEL A., DELEU J.
Belief Theory and Expert System.
Congrès Medinfo 1989 (à paraître).
- (BEUSCART 86) BEUSCART R., PARADINAS P., MOUSSU L., COMYN G. et le
Groupe SES
Le Systeme Expert Septicémie.
Proceeding "CIIAM 86 Artificial Intelligence", p 741-753, Ed HERMES
PARIS, 1986.
- (BEYTOUT 88) BEYTOUT J., BERNARD E., STAHL JP, BLANC D., BEUSCART C. et
le Groupe SES
Les septicémies mortelles : facteurs de létalité.
RICA, Paris, 4 Décembre 1986, Abstract 185/C 11.
- (BHATNAGAR 86) BHATNAGAR RK, KANAL L.N.
Handling uncertainty Information : a Review of Numéric and Non-Numeric
Methods.
Uncertainty in Artificial Intelligence. Machine Intelligence and Pattern
Recognition 4. Eds by Kanal LN and JF Lemmer 1986, pp 3-26.
- (BOOSE 87) BOOSE J.H.
Expertise transfer and complex problems : using AQUINAS is a knowledge
acquisition workbench for expert systems.
Int. Journ. of Man-Mach. Stud., 26, pp 3-28.
- (BUCHANAN 78) BUCHANAN BG, MITCHELL T.M.
Model directed learning of production rules.
Pattern directed inference Systems, Water man D.A. and Hayes Roth F.,
Eds Academic Press, New-York 1978.
- (BUCHANAN 84) BUCHANAN BG, SHORTLIFFE EH
Rule-Based Expert System. The Mycin Experiments of the Stanford Heuristic
Programming Project.
Addison-Wesley 1984.
- (BONNET 86) BONNET A, HATON JP and coll
Systèmes Experts, vers la maîtrise technique.
Inter Editions.

- (BOY 88) BOY G., FALLER B., SALLANTIN J.
Acquisition et ratification des connaissances.
Proc PRC Gréco, Toulouse, Mars 1988, pp 321-356.
- (BROTTIER 86) BROTTIER-MANCINI E, ECHINARD E., RAGNAUD JM et le
Groupe SES
Les localisations septiques au cours des septicémies.
RICA, Paris, 4 Décembre 1986, Abstract 184/C 11.
- (CHAITMAN 79) Chaitman BR, WATERS D. and coll.
The importance of clinical subsets in interpreting maximal treadmill
exercise test results : the role of multiple-lead systems.
Circulation 1979, 59(3), p 560-570.
- (COM 88) Bulletin du Conseil de l'ordre des Médecins
12ème Assemblée Générale. 28, 1988.
- (COMBY 88) COMBY S., FLANDROIS JP, PAVE A.
A contribution to the study of the discriminant capability of an Expert
System. Application to clinical bacteriology.
Artificial Intelligence and Cognitive Sciences, Manchester University Press,
1988, pp 305-322.
- (COHEN 60) COHEN
A coefficient of Agreement for Nominal scales
Educational and Physiological Measurement, 20, pp 37-46.
- (DAVIS 79) DAVIS R.
Interactive transfert of Expertise : Acquisition of New Inferences Rules.
Art. Int., 12, 1979, pp 121-157.
- (DE DOMBAL 74) DE DOMBAL FT, LEAPER DJ and coll
Human and computer aided diagnosis of abdominal pain : further reports
with emphasis on performance of clinicians.
Br. Med. J., 1974, 4, pp 376-380.
- (DELAHAYE 87) DELAHAYE JP
Outils logiques pour l'Intelligence Artificielle.
Eds Eyrolles.
- (DEMONGEOT 88) DEMONGEOT J., ROBERT C.
A study of different uncertainty coefficients used in Artificial Intelligence :
towards a natural definition of weights in semantic networks.
Artificial Intelligence and Cognitive Science, Manchester University Press,
1988, pp 177-178.

(DELEU 88a) DELEU J., BEUSCART R., BECQUART E., DUHAMEL A.
Comparaison of a Probabilistic Network and an Expert System.
Proceedings of IEEE/Engineering in Medicine and Biology Society, Nov. 13-16, 1988, NEW ORLEANS, pp 1335-1336.

(DELEU 88b) DELEU J., BEUSCART R., BECQUART E., DUHAMEL A., COMYN G.
Reseau Bayésien et Diagnostic Médical.
Actes Neuro-Nimes, 88, Nov. 15-17, 1988, pp 95-103.

(DIAMOND 79) DIAMOND GA, FORRESTER JS
Analysis of probability as an aid in the clinical Diagnosis of coronary artery disease.
N. Eng. J. Med., 300, 1979, p 1350-1358.

(DUHAMEL 88a) DUHAMEL A., BEUSCART C., BEUSCART R. et le Groupe SES
Le Système Expert Septicémies : optimisation des coefficients de vraisemblance grâce à l'analyse d'une base de cas.
Innovation et Technologie en Biologie et Médecine, Vol 9, n°5, 1988, p. 577, 584.

(DUHAMEL 88b) DUHAMEL A., BEUSCART R., DEMONGEOT J. et le Groupe SES.
Validating an expert system with a large epidemiological database.
Medical Informatics : computers in clinical Medicine, British Medical Informatics Society, 1988, p 15-19.

(DUHAMEL 88c) DUHAMEL A., BEUSCART R., DEMONGEOT J., MOUTON Y. et le Groupe SES.
SES (Septicemia Expert System) : Knowledge Validation from Data Analysis.
Proceedings of IEEE/Engineering in Medicine and Biology Society, Nov. 13-16, 1988, New Orleans, p 1400-1401.

(ERIKSSEN 77) ERIKSEN, FORFANG
Angina pectoris in presumably healthy middle-aged men.
Eur. J. Cardiol., 1977, 6 (4), pp 285-298.

(ESHELMAS 86) ESHELMAS L. Mc DERMOTT J.
Mole : a knowledge Acquisition tool that uses its head.
Proc AAAI-86, Philadelphia, 1986, pp 950-955.

(FARGEAS 88) FARGEAS X., FRYDMAN F.
Les Systèmes Experts en Médecine.
Hermès, Technologie de Pointe.

(FEIGENBAUM 84) FEIGENBAUM E, BARR A.
The Hard book of Artificial Intelligence.
Vol. 1

(FEIGENBAUM 85) FEIGENBAUM E., CORDUCK P.
La cinquième génération, le pari de l'Intelligence Artificielle à l'aube du
21ème siècle.
Inter-Editions, 1984.

(FARRENY 85) FARRENY H.
Les Systèmes Experts : Principes et exemples.
Techniques Avancées de l'Informatique, Cepadues Editions.

(FERRAND 84) FERRAND P.
SESAM : an explanatory medical aid system.
ECAI, 1984, Pise, Italy.

(FIESCHI 83) FIESCHI M.
Sphinx : un système d'aide à la Décision Médicale.
Thèse Doctorat Etat en Biologie Humaine, Marseille, 1983.

(FIESCHI 87) FIESCHI M.
Etude de mise à disposition du praticien généraliste d'un Système Expert
d'Aide à la Thérapeutique du Diabète : étude de faisabilité.
Rapport de Contrat ADI 84747, 1987.

(FIESCHI 88) FIESCHI M.
Conceptualisation, représentation et utilisation des connaissances dans les
Systèmes Experts Médicaux.
Systèmes Experts en Médecine, Joly, EC2.

(FIESCHI 84b) FIESCHI M.
Intelligence Artificielle en Médecine : des Systèmes Experts.
Masson 1984.

(FIESCHI 84a) FIESCHI M., JOUBERT M.
Some reflections on the Evaluation of Expert System in Medicine.
Working paper, MIT, Boston USA

(FOUCARD 85) FOUCARD T.
Analyse Factorielle. Programmation sur micro-ordinateurs.
Masson, 2eme ed., 1985.

(FOX 77) FOX J.
Medical computing and the user.
Int. J. of Man-Machine Studies, 77.

(FOX 85) FOX J., MYERS CD and coll
Knowledge acquisition for Expert Systems : experience in Leukaemia
diagnosis.
Method Inf. Med., Vol 24, pp 65-72.

(FRYBACK 78) FRYBACK D.G.

Bayes theorem and Conditional Nonindependence of Data in Medical Diagnosis.

Computers and Biomedical Research, 11, pp 423-434, reprinted in Computer-Assisted Decision Making, Springer-Verlag, pp 183-195.

(GASCUEL 81) GASCUEL O.

Un Système Expert dans le domaine médical.

Thèse 3ème cycle, Univ. P. et M. Curie, nov. 81.

(GASCHNING 83) GASCHNIG J., KLAHR P., POPLE H. and coll.

Evaluation of Expert Systems : issues and cases studies.

Hayes-Roth F., Waterman DA and Lenat D (Eds). Building Expert Systems, Addison - Wesley Company.

(GINSBERG 88) GINSBERG A., WEISS S.M. and coll

Automatic knowledge base refinement for classification systems.

Artificial Intelligence, 35, pp 197-226.

(GOLDMAN 88) GOLDMAN L. and others

A computer protocol to Predict Myocardial infarction in Emergency Department Patients with chest Pain.

NJM Mars 88, vol 318, n°13, p 797-803.

(GONDRAN 85) GONDRAN M.

Introduction aux Systèmes Experts.

Eds Eyrolles, Paris 1985.

(GOULET 86) GOULET V.

Les hémocultures en 1984. Enquête du Laboratoire National de la Santé.

Bull. Epid. Hebdo., 31, pp 121-122.

(GOUVERNET 85) GOUVERNET J., AYME S.

Gendiag : Système d'Aide au Diagnostic en Génétique Médicale.

Médecine Informatique, n° 15-16, pp 7-9.

(GREENBERG 84) GREENBERG PS, ELLESTAD MH and coll

Comparison of the Multivariate Analysis and Codenza Systems for Determination of the Probability of coronary Artery disease.

Am. J. Cardiol. 1984, 53, p 493-496.

(GREMY 87) GREMY F.

Informatique Médicale. Introduction à la méthodologie en Médecine et Santé Publique.

Médecine-Sciences, Flammarion, 1987.

- (GROSOF 86) GROSOF BN
 Evidential Confirmation as transformed probability on the duality of Priors and updates.
 Uncertainty in Artificial Intelligence. Machine Intelligence and Pattern Recognition 4. Eds by Kanal LN and JF Lemmer, 1986, pp 153-166.
- (HAYWARD 87) HAYWARD S.A.
 Structured Analysis of Knowledge.
 Int. J. Man-Machine Studies, 26, pp 487-498.
- (HECKERMAN 86) HECKERMAN D.
 Probabilistic Interpretation for Mycin's certainty factors.
 Uncertainty in Artificial Intelligence. Machine Intelligence and Pattern Recognition 4. Eds by Kanal LN and JF Lemmer 1986, pp 167-196.
- (HORVITZ 86) HORVITZ E., HECKERMAN D.
 The inconsistent use of measures of certainty in Artificial Intelligence Research.
 Uncertainty in Artificial Intelligence. Machine Intelligence and Pattern Recognition 4. Eds by Kanal LN and JF Lemmer 1986, pp 137-151.
- (HORWITZ 74) HORWITZ L.D.
 The diagnostic significance of anginal symptoms.
 JAMA 1974, 229(9), pp 1196-1199.
- (JOLY 88) JOLY H.
 Systèmes Experts en Médecine.
 Ed. EC2, 1988.
- (JULIEN 83) JULIEN JL, CACHERA JP and coll
 Angine de poitrine.
 Encycl. Med. Chir. Paris Coeur-Vaisseaux, 1983, 2, 11030 A10 A20 A30.
- (KAHN 85) KAHN G., NOWLAN S. and coll
 More : an intelligence knowledge acquisition tool.
 Proc. of 9th. Point Conference on Art. Intell., Los Angeles, 1985.
- (KANNEL 76) KANNEL WB, Mc GEE S. and coll
 A General cardiovascular risk profile : the Framingham Study.
 Am. J. of cardiol., 38 : 46, 1976.
- (KODRATOFF 88) KODRATOFF Y.
 Introduction to learning.
 Edition Pitman, 1988.

- (KULIKOWSKI 80) KULIKOWSKI G.A.
Artificial Intelligence methods and Systems for Medical Consultation.
IEEE Trans. Patt. Anal. Machine Intell., PAMI, 2, pp 464-475.
- (LACUT 86) LACUT JY, ARMANGAUD M., JANBON F. et le Groupe SES
Epidémiologie clinique et traitement des septicémies nosocomiales.
RICAI, Paris, 4 Décembre 1986, Abstract 183/C 11.
- (LANDAU 87) LANDAU JA, NORWICH KH, EVANS SJ and coll
An error correding protocol for Medical Expert Systems.
J. of Man-Machines Studies, 1987, 26, pp 617-625.
- (LAPLANCHE 87) LAPLANCHE A., COM-NOUGHE G., FLAMENT R.
But et principe de l'essai comparatif.
Méthodes statistiques appliquées à la Recherche clinique.
Flammarion, 1987, pp 1-14.
- (LAURIERE 82) LAURIERE J.L.
Représentation et utilisation des connaissances.
TPSI 82, Vol. 1, n°2.
- (LEBART 77) LEBART L., MORINEAU A., TABARD N.
Techniques de la Description Statistique. Méthodes et logiciels pour
l'analyse des grands tableaux.
Dunod 1977.
- (LEDLEY 59) LEDLEY RS, LUSTED LB
Reasoning Foundations of Medical Diagnosis.
Reprinted in Computer-Assisted-Medical-Decision-Making, Springer-Verlag,
Vol. 1, p 47-79.
- (MAC CABE 83) MAC CABE W.R. and coll.
Pathophysiology of bactériemia.
Am. J. Med. 75, pp 7-18.
- (MAC GOVAN 75) MAC GOVAN J.E. and coll
Bacteremia at Boston City Hospital : occurence and mortality during 12
selected years (1935-1972) with special reference to Hospital Acquired
cases.
J. of Inf. diseases, 132, n°3, pp 316-335.
- (MAC GREGOR 72) MAC GREGOR R.R. and coll
Evaluation of positive blood cultures.
Arch. Int. Med., 130, pp 84-87.

(MANUEL 85) MANUEL C.

Validation du Système Expert Sphinx dans son application a la thérapeutique du diabète. Etude sur 100 dossiers.

Thèse de Doctorat en Médecine, Marseille 85.

(MASSARI 86) MASSARI Ph. BREUX JP, FROTTIER J. et le Groupe SES

Les septicémies a porte d'entrée urinaire : épidémiologie, traitement antibiotique de première intention.

RICAI, Paris, 4 Décembre 1986, Abstract 182/C 11.

(MITCHELL 82) MITCHELL TM

Generalization as search.

Art. Intell., 18, 1982, pp 203-226.

(MODAI 84) MODAI J.

Qu'est-ce qu'une septicémie ?

J. Med. Infect., 14, n°10, pp 456-459.

(MOUTON 88) MOUTON Y.

Consensus, Systèmes Experts et Aide à la prescription d'antibiothérapie d'attente.

Réanimation-Soins Intensifs-Médecine d'Urgence, 1988, 4, n°2.

(MOUTON 87) MOUTON Y., DEBOSKER Y.

Le point actuel sur les antibiotiques.

Eds Laboratoires Bristol, 1987.

(NAKACHE 80) NAKACHE J.P.

Méthodes de discrimination sur variables de nature quelconque. Théorie et Pratique.

Thèse de Doctorat d'Etat de Mathématiques, Paris 1980.

(NILSSON 80) NILSSON N.

Principles of Artificial Intelligence.

Tioga, Palo Alto, Ca. 1980.

(PEARL 86) PEARL J.

Fusion, propagation and structuring in belief networks.

In : Artificial Intelligence, 29, pp 241-288.

(PEARL 87) PEARL J.

Distributed revision of composite beliefs.

In : Artificial Intelligence, 33, pp 173-215.

(PEARL 88) PEARL J.

Embracing causality in default reasoning.

In : Artificial Intelligence, 35, pp 259-271.

- (PEARL 88) PEARL J.
Probabilistic reasoning on Intelligent Systems : network of plausible inference..
Morgan-Kaufmann, 1988.
- (POLITAKIS 84) POLITAKIS P., WEISS S.
Using Empirical Analysis to Refine Expert System Knowledge Bases.
Art. Intell., 22, 1984, p. 23-48.
- (PORTIER 87) PORTIER H., MARCHOU B., JANBON F. and coll
French multicentric Septicemia Expert System (SES) : two years epidemiological survey.
ICAAC, New-York, Oct 87, abstract n° 132.
- (PRADE 85) PRADE H.
A combinational approach to approximate and plausible reasoning with applications to Expert Systems.
IEEE, trans on PAMI, vol 7, n°3, May 85, pp 260-283.
- (QUAGLINI 87) QUAGLINI S., STEFANELLI M, BAROSI G. and coll.
Evaluating the performance of Anemia.
Lectures notes in Medical Informatics, AIME 87.
- (REGGIA 85) REGGIA JA, RUHREN S.
An overview of Methods for Computer-Assisted-Decision-Making.
Computer-Assisted Medical Decision Making. Springer-Verlag, pp 3 - 45.
- (RIALLE 87) RIALLE V.
Aide au diagnostic et à l'apprentissage dans un domaine incertain, incomplet et évolutif.
These de l'USMG, 1987.
- (RICH 85) RICH E.
Artificial Intelligence.
International Student Edition.
- (ROBERT 89) ROBERT C.
Analyse des données - applications médicales.
Flammarion, à paraître.
- (ROBERT 88a) ROBERT C., DUHAMEL A., RIAILLE V., VILA A.
Using clinical datafiles in building expert systems.
Artificial Intelligence and Cognitive Sciences, Manchester University Press, 1988, p. 387, 393.
- (ROBERT 88b), ROBERT C.
Observable Network, a Markov field approach for probabilistic reasoning.
Soumis à IEEE.

(ROBERT 88c) ROBERT C. FRANCOIS P. and coll.
Multivariate Analysis : a Tool in Artificial Intelligence.
Medical Informatics : computers in clinical Medicine, British Medical
Informatics Society, 1988, pp 51-56.

(ROBERT 86) ROBERT C., ZARSKI JP, DEMONGEOT J.
Utilisation de la notion de score comme outil décisionnel en Médecine.
Proposition pour l'aide à la fabrication automatique de règles de
production.
Proc. IA-Biomed, Montpellier, 86, pp 234-243.

(ROMEDER 73) ROMEDER J.M.
Méthodes et programmes d'Analyse discriminante.
Dunod 1973.

(ROTTERDAM 86) ROTTERDAM EP, DEVRIES ROBBE PF
A critical comment of certainty factors.
Proc AI, Biomed 86, Montpellier pp 245-248.

(ROUE 86) ROUE R., SCHMIT JL, MARTIN D et le Groupe SES
Epidémiologie des septicémies à porte d'entrée pulmonaire.
RICAI, Paris, 4 Décembre 1986, Abstract 181/C 11.

(ROUSSEL 88) ROUSSEL Ph, DUHAMEL A., ANCEAUX F., BEUSCART R.
Génération de Langage naturel : GLOSE.
Forum jeunes chercheurs, Paris 26-27 Mai 1988.

(SAPORTA 78) SAPORTE G.
Théories et méthodes de la statistique.
Edition Technip, 1978.

(SAMPSON 71) SAMPSON J.J., CHEITLESS M.D.
Physiopathologie et diagnostic différentiel de la douleur cardiaque.
Acquis. Nouv. en Path. Cardio-vasc., 1971, N°3, p 555-577.

(SEBAG 87) SEBAG M., SCHOENAUER M.
Réduction, Extrapolation et Génération des coefficients de vraisemblance, à
partir de règles ou d'exemples.
Ecole Polytechnique n° 172, 1987.

(SHAFER 76) SHAFER G.
A mathematical theory of Evidence.
Princeton university Press, 1976.

(SHAMSOLMAALI A., CARSON ER and coll
Combining a Numerical Simulation with a knowledge based System.
Medical Informatics : computers in clinical Medicine. British Medical
Informatics Society, 1988, pp 103-108.

(SHAPIRO 77) SHAPIRO AR
The evalude of clinical predictions : A Method and Initial Application.
N.JM, 296 pp 1509-1514.

(SHORTLIFFE 75) SHORTLIFFE EH, BUSHANAN BG
A modif of Inexact reasoning in Medicine.
Mathematical Biosciences, 23, pp 251-379.

(SHORTLIFFE 84) SHORTLIFFE EH, DAVIS R., AXLINE SG and coll.
Computer based consultations in clinical therapeutics : the Mycin System.
Computer and Biomed., vol. 8, pp 303-320.

(SOULA 87) SOULA G, THIRION X and coll
Une validation multicentrique du système Protes.
Avignon 87, EC2, pp 239-255.

(SPIEGELHALTER 86) SPIEGELHALTER DJ
A statistical view of uncertainty in Expert Systems.
Artificial Intelligence and Statistics, Ed. A. Gale, pp 17-56, Addisson Wesley.

(SPIEGELHALTER 83) SPIEGELHALTER DJ
Evaluation of a clinical decision aids with an application to a system for
dysleptia.
Statistics in Med., vol 2, pp 207-216.

(STHAL 87) STHAL J.P., BEUSCART C., DUHAMEL A. et le Groupe SES.
Multivariate analysis of 450 septicemias.
Third European Congress of Clinical Microbiology Interdisciplinary Meeting,
The Hague, 1987, abstract 321.

(SVANBOM 79) SVANBOM M.
Septicemia I. A. projective Study on Etiology underlying factors ans sources
of infectious.
Scand. J. Infect. dis. 11, pp 187-198.

(SWARTOUT 81) SWARTOUT W.R.
Producing explanations and justification of expert consuting programs.
MIT, Janv. 81.

(SZOLOVITS 82) SZOLOVITS P.
Artificial Intelligence in Medicine.
AAAS, Selected Symposium Series, Westview Press, Inc., Colorado 1982.

- (TEACH 81) TEACH R., SHORTLIFFE E.
An analysis of Physicians attitudes regarding computer-based- clinical
consultation system.
Computers and Biomedical research 14, pp 542-558.
- (VAN MELLE 79) VAN MELLE W.
A domain independant production rule system for consultation programs.
Proc. Sixth int. Joint conf. on Art. Intell., 1979, 923-925.
- (VIGNON 86) VIGNON C.
Intérêt d'un questionnaire dans l'évaluation de la maladie coronarienne.
Thèse doctorat en Médecine, Lille 1986.
- (VOYER 87) VOYER R.
Moteurs de Systèmes Experts.
Ed. Eyrolles, 1987.
- (WARNER 61) WARNER H.R., TORONTO A.F. and coll
A Mathematical Approach to Medical Diagnosis : Application to Congenital
Heart Disease.
Journal of Am-Med. Ass., 22, 1961, vol 177, pp 177-183.
- (WASSON 85) WASSON JH, SOX HC, NEFF RK and coll
Clinical prediction rules : Application and methodological standards.
NJM. Vol. 313, pp 793-799.
- (WEBER 88) WEBER JE, BORTELS PH
Performance evaluation of an Expert System using rescaled certainty
factors.
Proc IEEE 88, 1371.
- (WEINSTEIN 83a) WEINSTEIN M.P. and coll
The clinical significance of positive blood cultures. Part I.
In : Rev. of Inf. dis., 5, n°1, pp 35-53.
- (WEINSTEIN 83b) WEINSTEIN M.P. and coll
The clinical significance of positive blood cultures. Part II.
In : Rev. Of Inf. dis., 5n n°1, pp 54-70.
- (WEINTRAUB 85) WEINTRAUB WS, BARR-ALDERFER VA and coll
A Sequential approach to the diagnosis of coronary artery disease using
multivariate analysis.
Am. Heart. J., 1985, 109, pp 999-1005.
- (WEISS 79) WEISS S., KULIKOWSKI C.A.
Expert : a system for Developing Consultation Models.
Proc. of Sixth Int. Joint Conference on Art. Intelligence.

(WEISS 85) WEISS S., KULIKOWSKI C.A.
A partical guide to Designing Expert Systems.
Rowman and Allanheld.

(WYATT 87) WYATT J.
The evaluation of clinical decision support systems. A discussion of the
methodology uses in the ACORN project.
Lectures notes in Medical Informatics, AIME 87, p 15-24.

(YU 79) YU VL, FOGAN LM, WRAITH SM and coll
Antimicrobial selection by computer : a blind evaluation by infectious
disease expert.
JAMA, vol. 242, pp 1279-1282.